Molecular Genetics of Arrhythmogenic Right Ventricular Cardiomyopathy in South Africa

Janine Blanckenberg

(née du Preez)

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Medicine

UNIVERSITY OF CAPE TOWN

March 2011

DECLARATION

I, Janine Blanckenberg (née du Preez), hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: ……

Date: 17 March 2011

ACKNOWLEDGEMENTS

The members of family ACM 2 who participated in this project.

My supervisor, Prof. Bongani Mayosi, for his input and enthusiasm throughout this project.

Prof. Hanlie Moolman-Smook for her valuable input while co-supervising part of the project.

Michelle Parker for assistance with microsatellite marker genotyping and for reading and providing feedback on Chapter 2.

Dr Nicki Tiffin for help with prioritisation of the genes in the ARVC-6 locus and for reading and providing feedback on Chapter 3.

Dr George Rebello for teaching me how to use the annotation program, ANNOTV9, which he designed.

Dr Gasna Shaboodien for assistance with mutation screening.

The Central Analytical Facility, Stellenbosch University for assistance with sequencing.

Prof. Peter Schwartz and the members of the Molecular Cardiology Laboratory in Italy, especially Dr Roberto Insolia and Dr Matteo Pedrazzini, for their support during the DHPLC mutation screening.

Maryam Fish for help with High Resolution Melt (HRM).

Jo McBride for Affymetrix® Genome-Wide Human SNP Array 6.0 genotyping.

Dr Judit Kumuthini and Kusha Kalideen for analysis of the Affymetrix® Genome-Wide Human SNP Array 6.0 genotyping data. Dr Judit Kumuthini and Dr Armin Deffur for reading and providing feedback on Chapter 5.

Dr Ian Tietjen for Illumina HumanCytoSNP-12 BeadChip genotyping and haplotype analysis of the chromosome 4 region.

Marie-Pierre Dubé for analysis of the Illumina HumanCytoSNP-12 BeadChip genotyping data.

Lisa Roberts, Alvera Vorster, Dr Gemma Carvill and Michelle Parker for their advice.

The following organisations for financial assistance:
DAAD
Department of Medicine, University of Cape Town
Ernst and Ethel Eriksen Trust
South African Heart Association
The South African National Research Foundation
UCT postgraduate funding office:
Glickman/Elliot Scholarship
KW Johnston Bequest
Marion Beatrice Waddel Bursary

My parents for their love, support and encouragement throughout my studies.

My husband, Rory, for all his love, support, encouragement and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

^{o}C	= degrees Celsius
µg	= microgram
µl	= microlitre
µM	= micromolar
A	= adenine
ACBD7	= acyl-Coenzyme A binding domain containing 7
AD	= autosomal dominant
APC	= Anterior Polar Cataract
AR	= autosomal recessive
ARMETL1	= arginine-rich, mutated in early stage tumors-like 1
ARVC	= arrhythmogenic right ventricular cardiomyopathy
BIND	= Biomolecular Interaction Network Database
BLAST	= Basic Local Alignment Search Tool
bp	= base pairs
C	= cytosine
C1QL3	= complement component 1, q subcomponent-like 3
C10orf38	= chromosome 10 open reading frame 38
C10orf97	= chromosome 10 open reading frame 97
C10orf111	= chromosome 10 open reading frame 111
Ca^{2+}	= calcium
CAD	= coronary artery disease
CASSA	= Cardiac Arrhythmia Society of Southern Africa
CDS	= coding sequence
CGAP	= Cancer Genome Anatomy Project
CN	= copy number

CNV(s) = copy-number variation(s)

CPGR = Centre for Proteomic and Genomic Research

CPVT = catecholaminergic polymorphic ventricular tachycardia

CUGBP2 = CUGBP, Elav-like family member 2 (approved symbol now *CELF2*)

CVD(s) = cardiovascular disease(s)

DCLRE1C = DNA cross-link repair 1C

DCM = dilated cardiomyopathy

DDE = Dragon Disease Explorer

DES = desmin

DGP = Disease Gene Prediction

dH_2O = distilled water

DHPLC = denaturing high-performance liquid chromatography

DM = dense midline

DMSO = dimethyl sulfoxide

DNA = deoxyribonucleic acid

dNTPs = deoxynucleotide triphosphate

DSC2 = desmocollin-2

dsDNA = double stranded DNA

DSG2 = desmoglein-2

DSP = desmoplakin

EC = excitation-contraction

EDTA = ethylenediaminetetraacetic acid

EST = expressed sequence tag

Exo I = exonuclease I

F = forward primer

FAM107B = family with sequence similarity 107, member B

FRMD4A = FERM domain containing 4A

g = gram

G = guanine

gDNA = genomic DNA

GO = Gene Ontology

HCM = hypertrophic cardiomyopathy

HGNC = HUGO Gene Nomenclature Committee

HRM = High Resolution Melt

HSPA14 = heat shock 70kDa protein 14

HUGO = Human Genome Organisation

IBD = identical-by-descent

IBS = identical-by-state

ICD = implantable cardioverter-defibrillator

IDP = inner dense plaque

IF(s) = intermediate filament(s)

ISFC = International Society and Federation of Cardiology

ITGA8 = integrin alpha 8

JUP = plakoglobin

kb = kilo-base

kcal = kilocalories

KEGG = Kyoto Encyclopedia of Genes and Genomes

L = litre

LAMR1P6 = laminin receptor 1 pseudogene 6

lod = logarithm of the odds

LV = left ventricle

LVNC = left ventricular non-compaction

MAPD = median absolute pairwise difference

MEIG1 = meiosis expressed gene 1 homolog (mouse)

mg = milligram

$MgCl_2$ = magnesium chloride

ml = milliliter

mM = millimolar

MRI = magnetic resonance image

mRNA = messenger RNA

NCBI = National Center for Biotechnology Information

ng = nanogram

NGS = next-generation sequencing

nm = not mentioned

NMT2 = N-myristoyltransferase 2

ODP = outer dense plaque

OLAH = oleoyl-ACP hydrolase

OMIM = Online Mendelian Inheritance in Man

PCR = polymerase chain reaction

PKP2 = plakophilin-2

PM = plasma membrane

PS-DVB = polystyrene-divinylbenzene

PTER = phosphotriesterase related

PTPLA = protein tyrosine phosphatase-like (proline instead of catalytic arginine), member A

QC = quality control

R = reverse primer

RCM = restrictive cardiomyopathy

RFA = radiofrequency catheter ablation

RPP38 = ribonuclease P/MRP 38kDa subunit

RV = right ventricle

RyR2 = cardiac ryanodine receptor

SAGE = Serial Analysis of Gene Expression

SAP = Shrimp Alkaline Phosphatase

SCD = sudden cardiac death

SNP(s) = single nucleotide polymorphism(s)

ssDNA = single stranded DNA

SUV39H2 = suppressor of variegation 3-9 homolog 2 (Drosophila)

SV(s) = sequence variant(s)

T = thymine

Ta = annealing temperature

TBE = Tris Borate EDTA

TEAA = triethylammonium acetate

TGFβ3 = transforming growth factor-β3

Tm = melting temperature

TMEM43 = transmembrane protein 43

UCT = University of Cape Town

UTR = untranslated region

WHO = World Health Organisation

w/v = weight per volume

ABSTRACT

Background: Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a heritable disorder characterised by progressive degeneration of the right ventricular myocardium, arrhythmias and an increased risk of sudden death at a young age. Fourteen chromosomal loci have been linked to ARVC and nine disease genes have been identified. Linkage analysis of a South African family was previously performed at ARVC loci 1 to 6. ARVC loci 1 to 5 were excluded as disease loci in this family based on lack of evidence for linkage. However, a peak lod score of 2.93 was obtained for the ARVC-6 locus which is highly suggestive of linkage. Subsequently another locus (ARVC-7) and five ARVC disease genes (ARVC loci 8 to 12) have been reported. The aim of this project was to identify the disease gene that causes ARVC in this family.

Methods: Exclusion mapping was performed in this family at the ARVC-7 locus and for the five ARVC disease genes. The positional candidate genes in the ARVC-6 locus were prioritised for mutation screening using several existing bioinformatics tools and tissue expression data. Fifteen genes were screened for mutations, in two affected and two unaffected family members, by sequencing, High Resolution Melt (HRM) and Denaturing High Performance Liquid Chromatography (DHPLC). Copy-number analysis of one affected family member and whole-genome single nucleotide polymorphism (SNP) linkage analysis of 12 family members was performed.

Results: The ARVC-7 locus and the five genes were excluded from causing disease in this family. Mutation screening of 15 genes in the ARVC-6 locus identified 36 sequence variants (SVs). Twenty-nine SVs have been reported as SNPs on the NCBI SNP database and seven SVs were novel. Only one SV, c.1641A>G in *DCLRE1C*, was found in the two affected family members and was not present in the two unaffected family members screened. This single base substitution results in a synonymous amino acid change, the significance of which is unknown. No copy-number variation (CNV) was found in the ARVC-6 region. Whole-genome SNP linkage analysis identified a peak multipoint lod score of 2.79 on chromosome 4 and 10. The chromosome 10 -1lod interval is ~2.69Mb

and correlates with the ARVC-6 locus. Haplotype analysis of the chromosome 4 region identified a linkage interval of ~7.18Mb.

Conclusion: Comprehensive mutation screening of the ARVC-6 locus on chromosome 10 has not identified any overt disease-causing mutations. The synonynous substitution found in *DCLRE1C* requires further investigation. Whole-genome SNP linkage analysis identified a novel locus for ARVC on chromosome 4 which is likely to harbour the disease gene as it is the longest shared haplotype among affected family members.

1 INTRODUCTION

1.1 Cardiovascular Disease

Cardiovascular diseases (CVDs) are the number one cause of death in the world (WHO). Over 80% of CVD deaths occur in low- and middle-income countries (WHO). In South Africa, the National Burden of Disease Study in the year 2000 reported CVD as the second most important cause of death, accounting for 17% of all deaths (Bradshaw et al., 2003). The Heart of Soweto Study reported 1593 new cases of CVD were diagnosed in 2006 at the Chris Hani Baragwanath Hospital in Soweto, South Africa (Sliwa et al., 2008). The most common primary diagnosis in this group was heart failure.

1.2 Heart Failure

In developed countries, the crude incidence of heart failure is estimated to range from one to five cases per 1000 per year and the crude prevalence is estimated as three to 20 per 1000 (Mendez and Cowie, 2001). Coronary artery disease is the most common cause of heart failure in these countries. There are no published population-based incidence or prevalence studies of heart failure from developing countries. Hospital-based studies in Africa indicate that heart failure contributes to 3-7% of admissions due to CVD (Mayosi, 2007). In sub-Saharan Africa, heart failure is mainly due to non-ischaemic causes such as hypertension, rheumatic heart disease, pericarditis and cardiomyopathy (Mayosi, 2007).

1.3 Cardiomyopathy

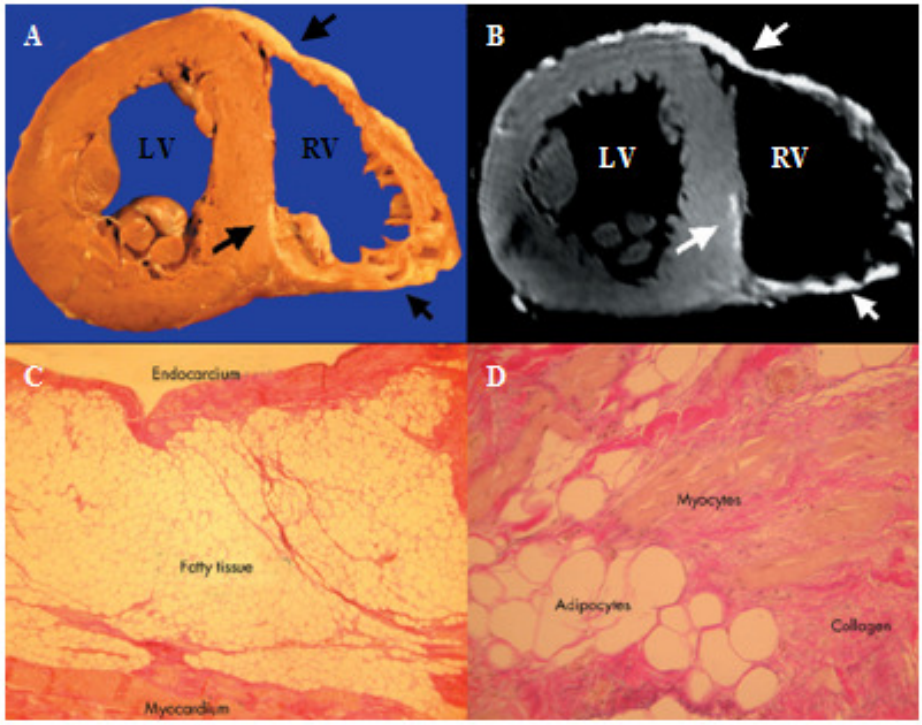
The European Society of Cardiology Working Group on Myocardial and Pericardial Diseases defines cardiomyopathy as "a myocardial disorder in which the heart muscle is structurally and functionally abnormal, in the absence of coronary artery disease, hypertension, valvular disease and congenital heart disease sufficient to cause the observed myocardial abnormality" (Elliott et al., 2008). There are five types of cardiomyopathy namely: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), restrictive cardiomyopathy (RCM), unclassified cardiomyopathy (e.g. left

ventricular non-compaction (LVNC)) and arrhythmogenic right ventricular cardiomyopathy (ARVC).

1.4 Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)

ARVC is characterised by fibrofatty replacement of the right ventricular myocardium (Marcus et al., 2010). Figure 1-1 shows the macroscopic structure, magnetic resonance image (MRI) and histopathology of a heart with ARVC. Four stages of disease progression have been documented: an early concealed phase, overt electrical disorder, isolated right heart failure and biventricular pump failure (Sen-Chowdhry et al., 2010). In the concealed phase, structural changes may be absent or subtle and confined to the right ventricular inflow tract, outflow tract or apex which is known as the "triangle of dysplasia" (Marcus et al., 2010). Individuals are often asymptomatic but may be at risk of sudden cardiac death (SCD), particularly during strenuous activity. The concealed phase is followed by an overt electrical disorder. Individuals present with symptomatic ventricular arrhythmia of RV origin and RV structural abnormalities are discernible on imaging (Sen-Chowdhry et al., 2010). Disease progression leads to global RV dysfunction and isolated right heart failure. Localised involvement of the left ventricle (LV) may occur. Progressive LV involvement leads to biventricular pump failure. Left–dominant and biventricular subtypes have also been recognised (Sen-Chowdhry et al., 2007b).

It is thought that ARVC may be a milder form of Uhl's anomaly. This disease is characterised by partial or complete absence of the right ventricular myocardium and it presents in infancy as heart failure. The pathogenesis of Uhl's anomaly has not been determined (Dokuparti et al., 2005).



(John et al., 2004)

Figure 1-1: A. A short axis slice of a heart with ARVC. Yellow streaks of fat are indicated with arrows. **B.** A MRI, of the same slice, which shows an increased signal intensity corresponding to the yellow streaks seen in *A*. **C&D.** A histopathology specimen taken from the RV showing large numbers of fat cells. LV = left ventricle, RV = right ventricle

1.4.1 Historical Background of ARVC

In 1961, Dalla-Volta and colleagues described an "auricularization of the right ventricular curve" in a patient with "primitive endocardial fibrosis" (Dalla Volta et al., 1961; Paul et al., 2003). In 1977, Fontaine and colleagues consolidated these findings and called it "right ventricular dysplasia" (i.e. an abnormality of development) (Michael et al., 2004; Paul et al., 2003). Once the cause of this disease was identified as progressive myocardial loss, the term "cardiomyopathy" was considered more appropriate than "dysplasia" (Thiene et al., 1991). In 1995, ARVC was added to list of cardiomyopathies by the WHO/ISFC Task Force (Richardson et al., 1996). The first case series of ARVC in South Africa was reported in 2000 by Munclinger and colleagues (Munclinger et al., 2000).

1.4.2 Epidemiology of ARVC

The exact prevalence of ARVC is unknown due to incomplete penetrance of the disease and under diagnosis by clinicians and pathologists (Paul et al., 2003; Wolfe and Corwin, 2005).

The prevalence of ARVC was determined as 1 in 1000 inhabitants served by the Academic Teaching Hospital in Quedlinburg, Germany (Peters et al., 2004). It was a retrospective study over a 5-year time period which revealed a much higher prevalence than that of other centres.

In Italy, preliminary studies showed ARVC to account for 20% of all sudden deaths in individuals under the age of 35 years and 27% of sudden deaths in young athletes (Corrado et al., 1990; Thiene et al., 1988). In a 21-year prospective cohort study, ARVC and coronary artery disease (CAD) were associated with the greatest risk of sudden death in Italian athletes (Corrado et al., 2003). In the USA, ARVC accounts for only 2% of sudden death in athletes while HCM is the most common cause of sudden death in athletes causing 30% of cases (Maron et al., 1996).

The regional variation in ARVC prevalence may be due to two reasons. First, the genetic predisposition, dietary pattern and environmental conditions of Italians are different from those of Americans (Williams and Chen, 2003). Second, in the 21-year Italian study, the examination of all hearts was performed by the same team of experienced cardiovascular pathologists (Corrado et al., 2003).

The scarcity of reports on ARVC in Africa is thought to be due to the lack of sophisticated cardiac electrophysiology facilities and expertise required for the diagnosis of the disease (Sliwa et al., 2005). A report from the ARVC Registry of South Africa suggests that ARVC occurs in all segments of the population and that its clinical features are similar to international findings (Watkins et al., 2009).

1.4.3 Diagnosis of ARVC

In 1994, an "International Task Force for ARVC" proposed standardised diagnostic criteria (McKenna et al., 1994). The diagnosis is based on major and minor criteria that encompass clinical, structural and functional features (Hodgkinson et al., 2005). A patient is diagnosed with ARVC if they have two major criteria, one major plus two minor criteria or four minor criteria from different diagnostic categories (McKenna et al., 1994). Subsequently, the Task Force Criteria have been modified to incorporate new knowledge and technology to improve sensitivity for early and familial disease (Marcus et al., 2010).

1.4.4 Treatment of ARVC

Treatment of ARVC is focused on controlling ventricular arrhythmias and preventing sudden death (Frias, 2005).

Patients with well-tolerated, non-life-threatening ventricular arrhythmias are treated with anti-arrhythmic drugs. Beta-blockers alone or in combination with sotalol or amiodarone may be used (Campbell, 2005). ARVC patients should avoid class Ia antiarrhythmic agents such as propafenone which has been associated with incessant ventricular tachycardia in these patients (Michael et al., 2004).

Patients with life-threatening ventricular arrhythmias are best treated with an implantable cardioverter-defibrillator (ICD). A study by Hodgkinson et al. demonstrated that ICD therapy improved survival of males. The effectiveness of ICD therapy in preventing sudden cardiac death in women was however not demonstrated (Hodgkinson et al., 2005). This supports the clinical observation of a milder disease expression and better long-term prognosis of ARVC in women (Wichter and Breithardt, 2005).

Patients with frequent ventricular tachycardias or ICD shocks despite optimal antiarrhythmic therapy are candidates for radiofrequency catheter ablation (RFA) (Herren et al., 2009).

In patients with severe right ventricular or biventricular failure, treatment consists of pharmacologic therapy for heart failure and heart transplantation (Frias, 2005; Michael et al., 2004).

1.4.5 Genetics of ARVC

ARVC is familial in 30 to 80% of cases and it is inherited in an autosomal dominant or an autosomal recessive manner (Gerull et al., 2004). Fourteen chromosomal loci have been linked to ARVC and nine disease genes have been identified, namely plakoglobin (*JUP*), cardiac ryanodine receptor (*RyR2*), desmoplakin (*DSP*), plakophilin-2 (*PKP2*), transforming growth factor-β3 (*TGFβ3*), desmoglein-2 (*DSG2*), desmocollin-2 (*DSC2*), transmembrane protein 43 (*TMEM43*) and desmin (*DES*). The most frequent mutations occur in desmoplakin, plakophilin-2 and desmoglein-2 (Herren et al., 2009). The chromosomal loci and disease genes are given in Table 1-1.

Table 1-1: Chromosomal loci and genes involved in ARVC

Type of ARVC	Reference	OMIM identifier	Chromosome locus	Mode of inheritance	Gene & OMIM identifier	Reference
ARVC-1	(Rampazzo et al., 1994)	107970	14q23-q24	AD	Transforming Growth Factor-β3 (*TGFβ3*) (190230)	(Beffagna et al., 2005)
ARVC-2	(Rampazzo et al., 1995)	600996	1q42-q43	AD	Cardiac Ryanodine Receptor (*RyR2*) (180902)	(Tiso et al., 2001)
ARVC-3	(Severini et al., 1996)	602086	14q12-q22	AD	None	-
ARVC-4	(Rampazzo et al., 1997)	602087	2q32.1-q32.3	AD	None	-
ARVC-5	(Ahmad et al., 1998)	604400	3p23	AD	Transmembrane Protein 43 (*TMEM43*) (612048)	(Merner et al., 2008)
ARVC-6	(Li et al., 2000a) (Matolweni et al., 2006)	604401	10p12-p14	AD	None	-
ARVC-7	(Melberg et al., 1999)	609160	10q22.3	AD	None	-
ARVC-8	(Rampazzo et al., 2002)	607450	6p24	AD AR	Desmoplakin (*DSP*) (125647)	(Rampazzo et al., 2002) (Alcalai et al., 2003)
ARVC-9	-	609040	12p11	AD AR	Plakophilin-2 (*PKP2*) (602861)	(Gerull et al., 2004) (Awad et al., 2006)
ARVC-10	-	610193	18q12.1	AD	Desmoglein-2 (*DSG2*) (125671)	(Pilichou et al., 2006)
ARVC-11	-	610476	18q12.1	AD	Desmocollin-2 (*DSC2*) (125645)	(Syrris et al., 2006) (Heuser et al., 2006)
ARVC-12	-	611528	17q21	AD	Plakoglobin (*JUP*) (173325)	(Asimaki et al., 2007)
Naxos disease	(Coonar et al., 1998)	601214	17q21	AR	Plakoglobin (*JUP*) (173325)	(McKoy et al., 2000)
ARVC/APC	(Frances et al., 1997)	115650	14q24-qter	AR	None	-
ARVC-13	-	nm	2q35	AD	Desmin (*DES*) (125660)	(van Tintelen et al., 2009)

Note: The types of ARVC (i.e. ARVC-1 to -13) are classified based on the different loci mapped for the condition (Dokuparti et al., 2005).

APC = Anterior Polar Cataract; OMIM = Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim/); AD = Autosomal Dominant; AR = Autosomal Recessive, nm = not mentioned

1.4.5.1 Desmosomal genes

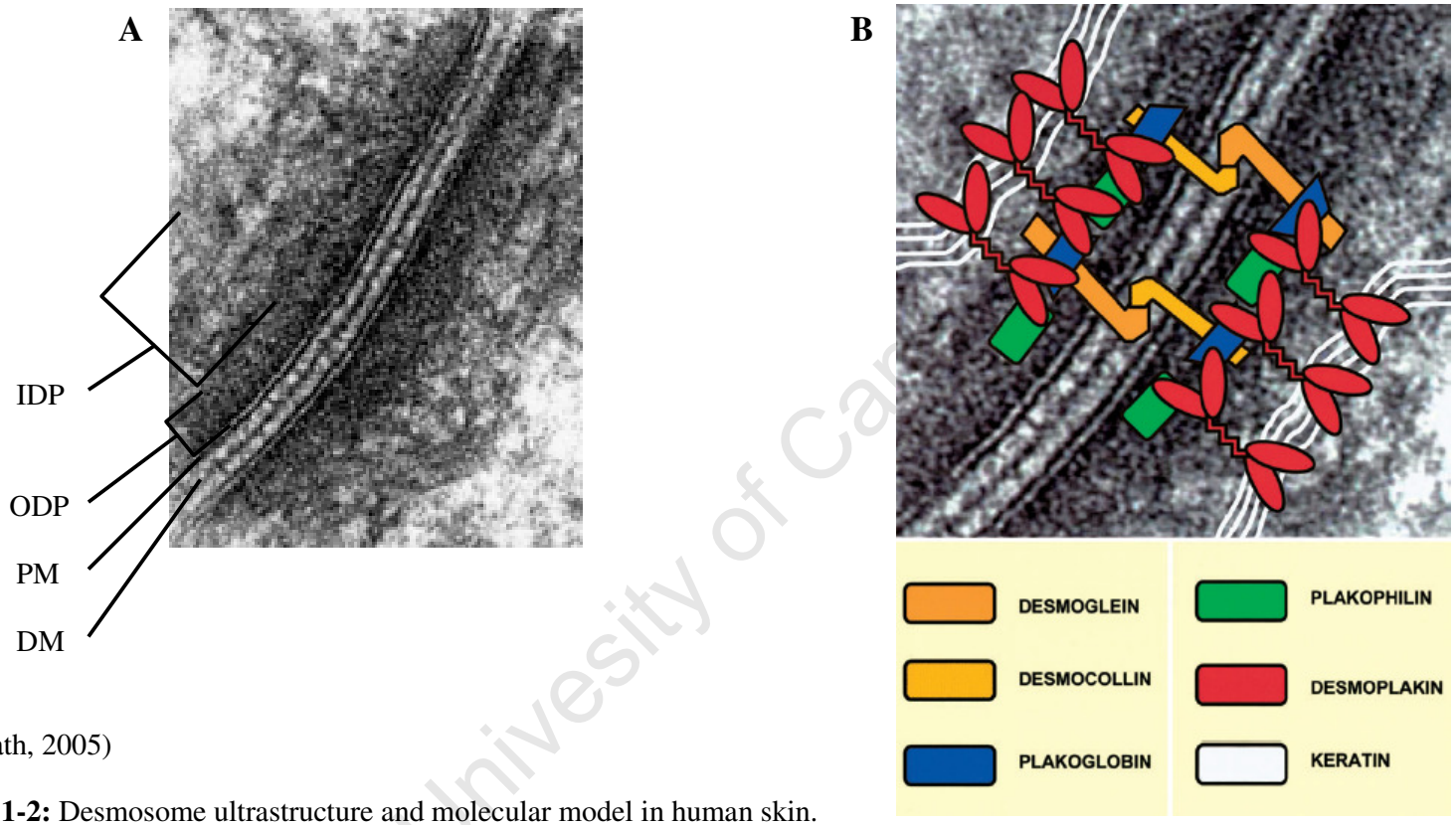
Five out of the nine ARVC genes, namely plakoglobin, desmoplakin, plakophilin-2, desmoglein-2 and desmocollin-2, encode cell junction proteins that are found in the desmosome. ARVC has therefore been called "a disease of the desmosome" (Gerull et al., 2004).

Desmosomes belong to a class of intercellular junctions called "anchoring junctions" (Yin and Green, 2004). They are found primarily in epithelial tissues but they are also present in the meninges, the dendritic reticulum cells of lymph node follicles and the myocardium. Desmosomes provide mechanical integrity to these tissues by anchoring intermediate filaments (IFs) to sites of strong adhesion (McGrath, 2005).

Desmosomes contain proteins from three distinct gene families, namely:

1. **cadherins:** desmogleins (DSG1-4) and desmocollins (DSC1-3)
2. **armadillo proteins:** plakoglobin, plakophilins (PKP1-3) and p0071 (also known as PKP4)
3. **plakins:** desmoplakins I and II, plectin, envoplakin, periplakin and epiplakin (McGrath, 2005; Yin and Green, 2004).

Figure 1-2 depicts the ultrastructure of the desmosome and the interaction of the major desmosomal proteins.



(McGrath, 2005)

Figure 1-2: Desmosome ultrastructure and molecular model in human skin.

A. An electron micrograph of a desmosome in human skin, showing the regions of the junction, namely: the inner dense plaque (IDP), the outer dense plaque (ODP), the plasma membrane (PM) and the dense midline (DM). **B.** A molecular model which depicts the relative organisation of the major desmosomal proteins.

1.4.5.1.1 Plakoglobin (*JUP*)

In 2000, plakoglobin was identified as the first ARVC disease gene. McKoy and colleagues reported a homozygous two base pair deletion in plakoglobin as the cause of Naxos disease (McKoy et al., 2000). Naxos disease was first described in 1986 in families originating from the Greek island of Naxos (Protonotarios et al., 1986). This disease is an autosomal recessive form of ARVC associated with thickening of the palms and soles (palmoplantar keratoderma) and woolly hair (Protonotarios et al., 2002). The skin and hair phenotype is fully expressed from early infancy, while the cardiac abnormalities are 100% penetrant by adolescence (Protonotarios et al., 2001). In 2007, the first dominant mutation in plakoglobin was identified in a German family with ARVC but with no cutaneous abnormalities (Asimaki et al., 2007).

1.4.5.1.2 Desmoplakin (*DSP*)

A genome-wide scan and mutation screening in a family with autosomal dominant ARVC identified desmoplakin as a ARVC disease gene (Rampazzo et al., 2002). A year later, a homozygous mutation in desmoplakin was identified in a family with autosomal recessive ARVC, pemphigous-like skin disorder and woolly hair (Alcalai et al., 2003).

1.4.5.1.3 Plakophilin-2 (*PKP2*)

Gerull and colleagues were the first to report plakophilin-2 as an ARVC disease gene (Gerull et al., 2004). They identified heterozygous mutations in 32 of 120 unrelated probands of western European descent (27%). Awad and colleagues reported the first case of recessive ARVC due to a homozygous cryptic splice mutation in *PKP2* (Awad et al., 2006). The patient had no significant cutaneous abnormalities.

1.4.5.1.4 Desmoglein-2 (*DSG2*)

Desmoglein-2 is the only desmoglein isoform expressed in cardiac myocytes. Pilichou and colleagues therefore screened this gene in 80 unrelated ARVC probands and identified nine heterozygous mutations in eight probands (10%) (Pilichou et al., 2006).

1.4.5.1.5 Desmocollin-2 (*DSC2*)

Syrris and colleagues were the first to report desmocollin-2 as an ARVC disease gene (Syrris et al., 2006). They screened 77 probands and identified two heterozygous mutations in 4 probands. A month later, Heuser and colleagues reported a mutation screen of *DSC2* in 88 probands. In one proband they identified a heterozygous transition that affects the conserved 3' splice-acceptor site of intron 5. Alternatively, a cryptic splice-acceptor site in exon 6 is used which causes a frameshift and a premature stop codon (Heuser et al., 2006).

Compound and double/digenic heterozygosity has also been identified in the desmosomal genes (Bauce et al., 2010; Xu et al., 2010).

1.4.5.2 Extradesmosomal genes

1.4.5.2.1 Cardiac Ryanodine Receptor (*RyR2*)

RyR2 is a calcium (Ca^{2+})-release channel found on the sarcoplasmic reticulum (an intracellular Ca^{2+} storage organelle). In the heart, Ca^{2+} regulates muscle contraction and electrical signals that determine the cardiac rhythm. RyR2 is required for excitation-contraction (EC) coupling and maintaining intracellular calcium homeostasis (Chelu et al., 2004; Marks, 2000; Tiso et al., 2001).

Mutations in *RyR2* cause ARVC-2 (Tiso et al., 2001) and catecholaminergic polymorphic ventricular tachycardia (CPVT) (Priori et al., 2001). CPVT is characterised by stress-induced, bidirectional ventricular tachycardia in the absence of structural heart disease (Laitinen et al., 2001; Priori et al., 2001). ARVC-2 differs clinically from the other types of ARVC due to the presence of effort-induced ventricular arrhythmias, high penetrance and 1:1 male:female ratio among affected patients (Tiso et al., 2001).

It has been hypothesised that mutations in *RyR2* cause ARVC-2 by altering the ability of the channel to remain closed which leads to Ca^{2+} leaking from the channel. The leaking of

Ca^{2+} will in turn lead to arrhythmias and an imbalance of intra-cellular Ca^{2+} can trigger apoptosis (Chelu et al., 2004; Dokuparti et al., 2005).

1.4.5.2.2 Transforming Growth Factor-β3 (*TGFβ3*)

In 2005 Beffagna and colleagues reported transforming growth factor-β3 (*TGFβ3*) as the disease gene involved in ARVC-1. TGFβ3 belongs to a group of regulatory cytokines which are important in development and tissue homeostasis. They identified a nucleotide substitution in the 5' untranslated region (UTR) of an ARVC-1 family and in the 3' UTR of an unrelated ARVC patient. Neither nucleotide change was found in 300 control subjects and *in vitro* expression assays showed mutated UTRs to be about 2.5-fold more active than wild-types (Beffagna et al., 2005).

In an editorial, Nattel and Schott mention a number of issues that need to be resolved before TGFβ3 overexpression can be accepted as the molecular basis of ARVC-1. Their first concern is that two families with linkage to ARVC-1 do not contain *TGFβ3* mutations. The *in vivo* overexpression of TGFβ3 needs to be confirmed in the mutation carriers and the biological connection between TGFβ3 overexpression and ARVC needs to be established. TGFβ3 is ubiquitously expressed yet mutations in this gene result in only a cardiac phenotype (Nattel and Schott, 2005).

1.4.5.2.3 Transmembrane Protein 43 (*TMEM43*)

A missense mutation was identified in *TMEM43* in fifteen unrelated ARVC families from the island of Newfoundland, Canada (Merner et al., 2008). The mutation occurs within a highly conserved transmembrane domain. TMEM43 contains a response element for an adipogenic transcription factor, PPARγ, which may explain the fibrofatty replacement of the myocardium. Merner and colleagues concluded that ARVC, due to this *TMEM43* mutation, is: "a lethal, fully penetrant, sex-influenced morbid disorder".

1.4.5.2.4 Desmin (*DES*)

Desmin is a major intermediate filament protein of skeletal and cardiac muscle. Van Tintelen and colleagues reported five Dutch families with an identical desmin mutation. An investigation of the clinical and pathologic characteristics of 27 patients identified two patients who fulfilled the diagnostic criteria for ARVC (van Tintelen et al., 2009).

1.5 ARVC research in South Africa

The ARVC Registry of South Africa was established in 2004 by the Cardiac Arrhythmia Society of Southern Africa (CASSA) (Latib et al., 2004). Suspected cases of ARVC are referred to the co-ordinating centre in the Cardiac Clinic at Groote Schuur Hospital, Cape Town. As of April 2009, 50 unrelated ARVC patients were enrolled in the registry (Watkins et al., 2009).

Genetic analysis was previously performed on the largest family enrolled in the registry (family ACM 2) (Matolweni et al., 2006). This family is of northern European descent and ARVC is inherited in an autosomal dominant manner. When genetic analysis began on this family there were only six known ARVC loci and no disease genes had been identified yet. In order to identify the disease-causing gene in this family, the family members were genotyped with microsatellite markers at the six known loci. ARVC loci 1 to 5 were excluded as disease loci and a peak lod score of 2.93 was obtained for the ARVC-6 locus. According to Matolweni and colleagues "lod score simulation analysis showed that 2.93 is the highest lod score that could be obtained for this pedigree" (Matolweni et al., 2006). Two positional candidate genes, integrin alpha 8 (*ITGA8*) and FERM domain containing 4 (*FRMD4A*), and a non-positional candidate gene, laminin receptor 1 pseudogene 6 (*LAMR1P6*), were screened for mutations. No disease-causing mutations were identified in these genes.

The current genetic research is focused on screening ARVC patients, enrolled in the registry, for mutations in the five desmosomal genes (i.e. plakophilin-2 (Watkins et al., 2009), desmoplakin (Fish, 2010), desmoglein-2, desmocollin-2 and plakoglobin).

1.6 The benefits of genetic elucidation of ARVC

Clinical diagnosis of ARVC can be difficult to make but genetic testing can aid diagnosis (Sen-Chowdhry et al., 2007a). The more is known about the genetics of this disease the more useful genetic testing will become as a diagnostic tool. If a mutation is identified in the proband, cascade screening of the family can be performed. This will identify the gene-carriers allowing targeting of clinical resources and avoiding lifelong clinical reassessment of extended families (Sen-Chowdhry et al., 2007a). Genetic counselling of patients is currently problematic due to incomplete penetrance and variable expressivity (Sen-Chowdhry et al., 2007a). Studies to determine mutation-specific cardiac features are therefore needed for accurate risk stratification (Merner et al., 2008). The identification of all ARVC disease genes will aid elucidation of the pathogenic pathway leading to the condition and identify targets for disease-modifying therapy (Dokuparti et al., 2005; Sen-Chowdhry et al., 2010).

1.7 The aim of this project

The aim of this project was to identify a novel disease gene that causes ARVC in the South African family of European descent (family ACM 2).

2 EXCLUSION MAPPING

2.1 Introduction

Linkage analysis of family ACM 2 was previously performed at six known ARVC loci (Matolweni et al., 2006). ARVC loci 1 to 5 were excluded as disease loci in this family based on lack of evidence for linkage. However, a peak lod score of 2.93 was obtained for the ARVC-6 locus which is highly suggestive of linkage. When this project was started, another locus (ARVC-7) and five ARVC disease genes (i.e. desmoplakin, plakophilin-2, desmoglein-2, desmocollin-2 and plakoglobin) had since been reported. The chromosomal positions of the loci are given in Table 2-1. The aim of this study was to exclude ARVC-7 and the five genes as the disease-causing loci for this family.

Table 2-1: ARVC loci and genes identified for exclusion mapping in family ACM 2

Type of ARVC	Chromosome	Gene identified
ARVC-7	10q22.3	None
ARVC-8	6p24	Desmoplakin (*DSP*)
ARVC-9	12p11	Plakophilin-2 (*PKP2*)
ARVC-10	18q12.1	Desmoglein-2 (*DSG2*)
ARVC-11	18q12.1	Desmocollin-2 (*DSC2*)
ARVC-12 and Naxos disease	17q21	Plakoglobin (*JUP*)

2.2 Methods

2.2.1 Patients

Twenty family members provided informed consent to participate in genetic studies to identify the causal mutation (Appendix A) and donated blood for DNA extraction. The clinical evaluation of the family members is described in Matolweni et al., 2006. The blood specimens and consent forms were sent to the molecular genetics laboratory of the Division of Human Genetics, University of Cape Town (UCT). Patient information was

captured onto the Human Genetics Laboratory Database and a laboratory reference code was assigned to each specimen to ensure confidentiality. DNA was extracted from lymphocytes by a research assistant using the PUREGENE™ DNA Isolation Kit (*Gentra Systems*, Adcock Ingram Scientific).

2.2.2 Polymerase Chain Reaction (PCR)

2.2.2.1 Primer Design

Three microsatellite markers were chosen to genotype the ARVC-7 locus as the disease gene has not yet been identified. These microsatellite markers were found to segregate with ARVC in the family in which the locus was mapped (Melberg et al., 1999). The Ensembl database (http://www.ensembl.org/index.html) was used to select a single published microsatellite marker within or in close proximity to each of the five disease genes. PCR primer sequences (Department of Molecular and Cell Biology, UCT) for the microsatellite markers were obtained from the UniSTS database on NCBI (http://www.ncbi.nlm.nih.gov/unists/) and are given in Appendix B (set one). Reverse e-PCR (http://www.ncbi.nlm.nih.gov/sutils/e-pcr) was used to check that the primers were specific for the amplification of the microsatellite markers. If the tandem repeat of a microsatellite marker was not indicated on NCBI, the Tandem Repeats Finder software (version 2.02) was used to identify the tandem repeat.

Three of the eight microsatellite markers were found to be uninformative for this family, namely D6S1547 (*DSP*), D12S1692 (*PKP2*) and AFM344ZD5 (*DSC2*). All CA and GT dinucleotide repeats, with a minimum of 20 repeats, were identified within 300 kb either side of the *DSP*, *PKP2* and *DSC2* genes. Two new microsatellite markers were designed for each of these three genes – one of which had to be within the gene if the published microsatellite marker was not located within the gene. Primers were designed using Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_results.cgi). In order to multiplex the six microsatellite markers (i.e. genotype the six microsatellite markers together), primers were chosen so that each microsatellite marker had a different PCR product size and the fluorescent tags were assigned based on PCR product size. The sequence of these primers (Department of Molecular and Cell Biology, UCT) is given in Appendix B (set two).

2.2.2.2 PCR Optimisation

PCR was optimised using the temperature gradient function on either the Labnet MultiGene Thermal Cycler (Whitehead Scientific) or the Px2 Thermal Cycler (*Thermo Electron Corporation*) (Appendix C). The PCR products were separated by gel electrophoresis through a 1% agarose gel and visualised using ethidium bromide (Appendix D). A volume of 5µl of PCR product was mixed with 3µl Green GoTaq^{R} Flexi Buffer (*Promega*, Whitehead Scientific) (Appendix E) and electrophoresed alongside 650ng 100bp DNA Ladder (*Promega*, Whitehead Scientific) (Appendix F). Four microsatellite markers (AFM344ZD5, D18S36, D10S2327 and D10S201) had small PCR product sizes (between 132bp and 230bp). These PCR products were separated by gel electrophoresis through a 3% MS-8 Agarose gel and visualised using ethidium bromide (Appendix D). Agarose MS-8 was used as it improves the resolution of small PCR products. A volume of 5µl of PCR product was mixed with 3µl sucrose solution (Appendix E) and electrophoresed alongside 0.5µg Low Molecular Weight DNA Ladder (*New England Biolabs*) (Appendix F). Sucrose was used instead of agarose loading dye as the bromophenol blue might have obscured the PCR product. The annealing temperature chosen for PCR of each microsatellite marker is given in Appendix B.

2.2.2.3 Amplification of Patient DNA

Microsatellite markers were amplified in family members using the optimised PCR conditions (Appendix C). A negative control, where distilled H_2O (dH_2O) replaced DNA, was included with PCRs set up to rule out contamination. PCR cycling was carried out on either the Labnet MultiGene Thermal Cycler (Whitehead Scientific) or the Px2 Thermal Cycler (*Thermo Electron Corporation*) at the cycling conditions given in Appendix C and the annealing temperatures in Appendix B.

2.2.3 Genotyping

Genotyping of microsatellite markers was performed on the ABI Prism® 3100 Genetic Analyzer (*Applied Biosystems*) (Appendix C) in the Division of Human Genetics, UCT and data was analysed using the GeneMapper version 3.0 software (*Applied Biosystems*).

Set one was genotyped in two multiplexes and a single multiplex was sufficient for set two.

Two-point Linkage Analysis, to test each microsatellite marker for linkage to ARVC, was performed using the LINKAGE program (version 5.10).

2.3 Results

Microsatellite marker genotypes for each family member are shown in Figure 2-1 to Figure 2-6. The lod scores for six recombination fractions (θ) are provided in Table 2-2 to Table 2-7. A lod score (Z) of 3.0 is the threshold for accepting linkage. Linkage can be rejected if $Z < -2.0$ and values of Z between -2 and +3 are inconclusive (Strachan and Read, 2004a). The recombination fraction is a measure of the genetic distance between two loci (in this case between the microsatellite marker and ARVC).

2.3.1 ARVC-7

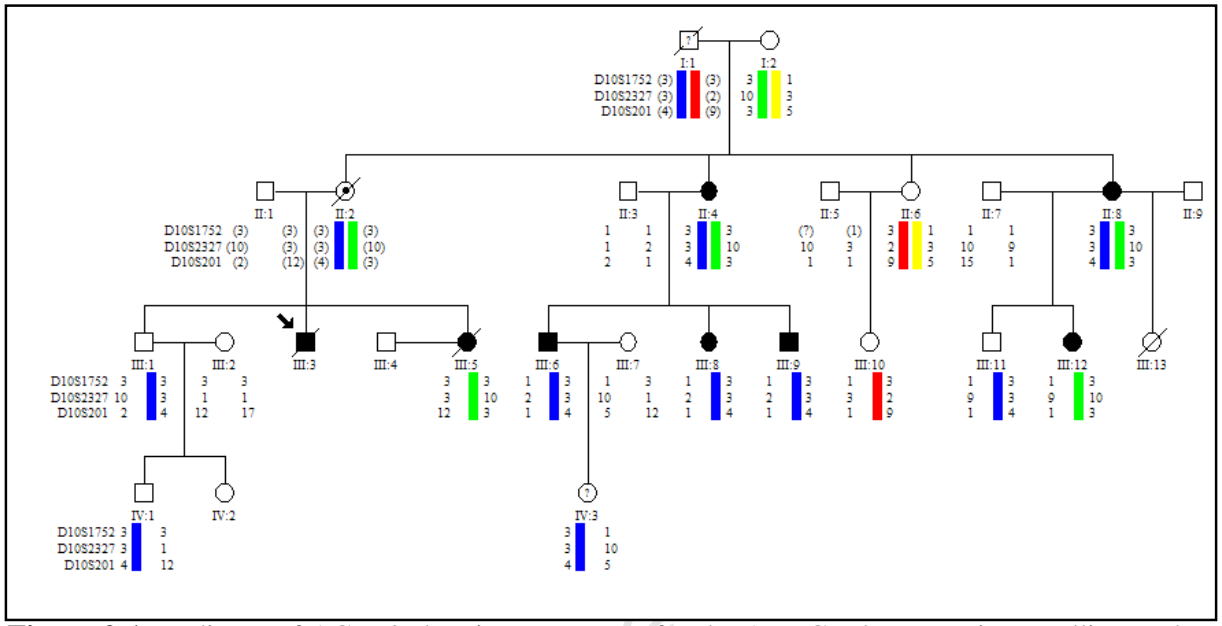


Figure 2-1: Pedigree of ACM 2 showing genotypes for the ARVC-7 locus. Microsatellite marker names are shown at the left of each generation. Genotypes in brackets are inferred.

Table 2-2: Two-point lod scores obtained for microsatellite markers in the ARVC-7 locus.

		Lod score					
Marker	**Chromosomal position (in bp)**	**θ = 0**	**θ = 0.1**	**θ = 0.2**	**θ = 0.3**	**θ = 0.4**	**θ = 0.5**
D10S1752	78,331,531 - 78,331,811	0.01	0.01	0	0	0	0
D10S2327	80,712,050 - 80,712,249	-43,429,448, 190,325,185, 656.55	0.13	0.20	0.13	0.03	0
D10S201	81,029,072 - 81,029,271	0	0	0	0	0	0

Haplotype 334 (represented in blue) is prominent in this family. It is present in five affected family members (II:4, II:8, III:6, III:8 and III:9). It is however not present in two affected family members (III:5 and III:12) and it is present in three unaffected family members (III:1, III:11 and IV:1). This haplotype is therefore not tracking with disease in the family. The lod score for D10S2327 at θ = 0 is less than -2 which excludes this locus as a disease locus for this family.

2.3.2 Desmoplakin

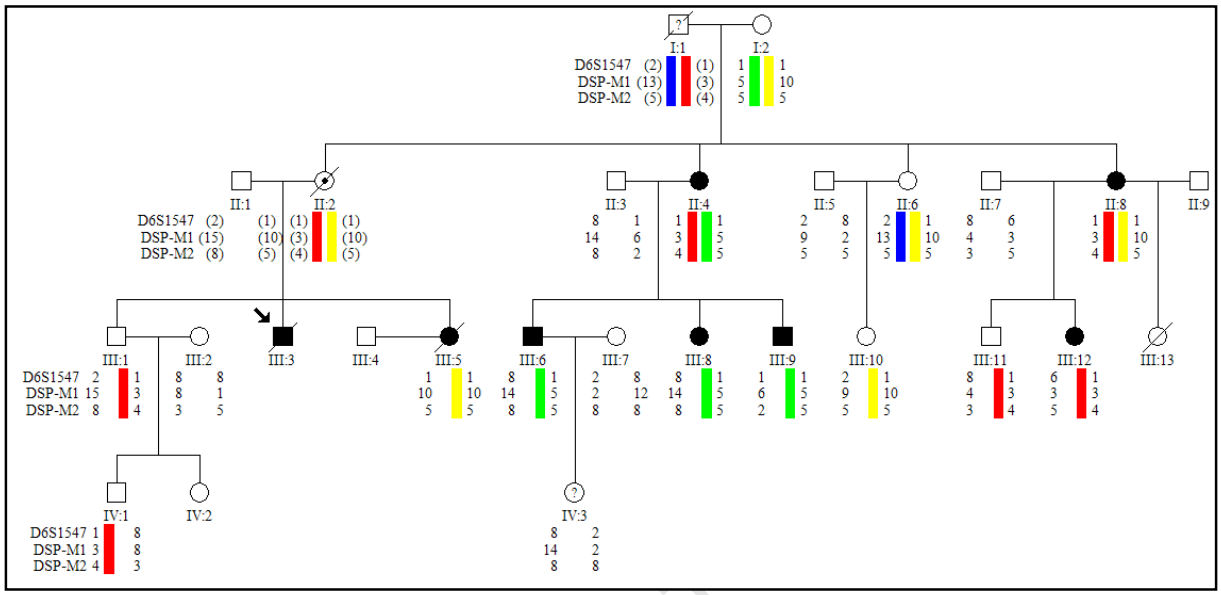


Figure 2-2: Pedigree of ACM 2 showing genotypes for the desmoplakin gene. Microsatellite marker names are shown at the left of each generation. Genotypes in brackets are inferred.

Table 2-3: Two-point lod scores obtained for microsatellite markers within and flanking the desmoplakin gene.

		Lod score					
Marker	**Chromosomal position (in bp)**	**θ = 0**	**θ = 0.1**	**θ = 0.2**	**θ = 0.3**	**θ = 0.4**	**θ = 0.5**
D6S1547	7,514,077 - 7,514,371	0.27	0.11	-0.03	-0.11	-0.10	0
DSP-M1	7,554,215 - 7,554,454 (in intron 1 of gene)	-4.14	-1.32	-0.80	-0.53	-0.30	0
DSP-M2	7,800,154 - 7,800,382	-0.86	-0.46	-0.47	-0.45	-0.29	0

Desmoplakin was excluded as a candidate gene as no haplotype is tracking with disease and a lod score less than -2 was observed at θ = 0 for DSP-M1 (which lies within the gene).

2.3.3 Plakophilin-2

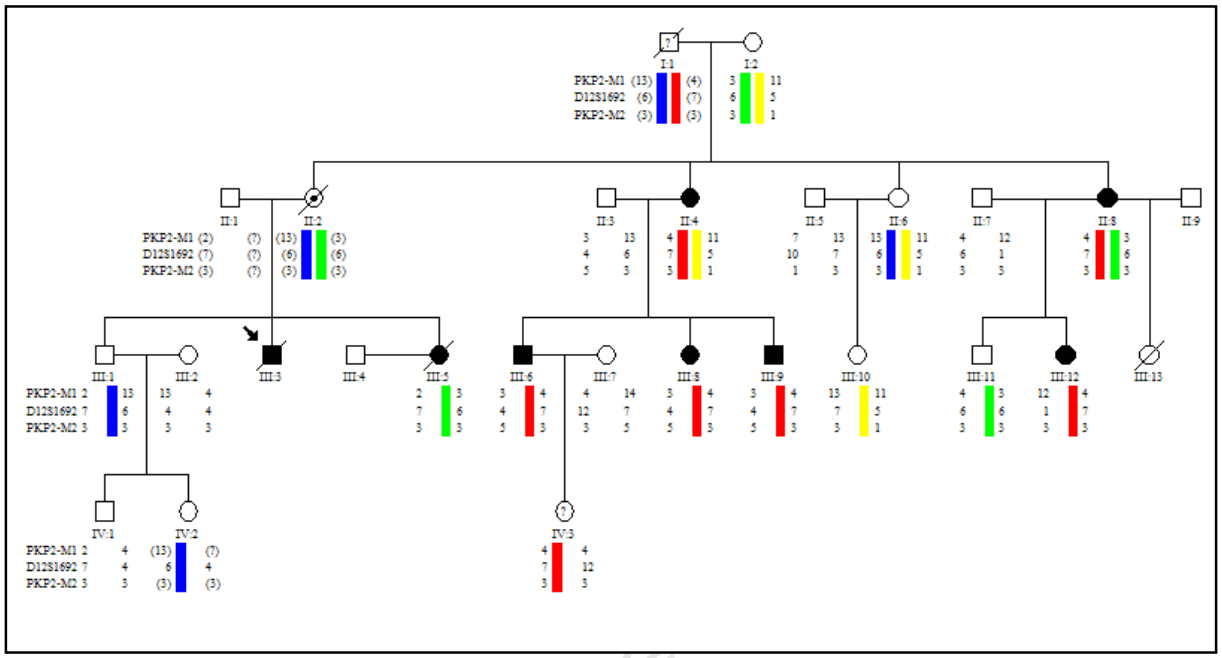


Figure 2-3: Pedigree of ACM 2 showing genotypes for the plakophilin-2 gene. Microsatellite marker names are shown at the left of each generation. Genotypes in brackets are inferred.

Table 2-4: Two-point lod scores obtained for microsatellite markers within and flanking the plakophilin-2 gene.

		Lod score					
Marker	**Chromosomal position (in bp)**	**θ = 0**	**θ = 0.1**	**θ = 0.2**	**θ = 0.3**	**θ = 0.4**	**θ = 0.5**
PKP2-M1	33,085,390 - 33,085,570	0.02	-0.03	0.19	0.25	0.18	0
D12S1692	32,987,723 - 32,987,971 (in intron 7 of gene)	1.05	1.32	1.12	0.79	0.39	0
PKP2-M2	32,849,128 - 32,849,468	0.77	0.66	0.51	0.35	0.17	0

Haplotype 473 (represented in red) is present in all the affected family members genotyped except for family member III:5. It is therefore unlikely that plakophilin-2 is the disease-causing gene in this family. The lod score values do not support linkage of this locus with disease in this family.

2.3.4 Desmoglein-2

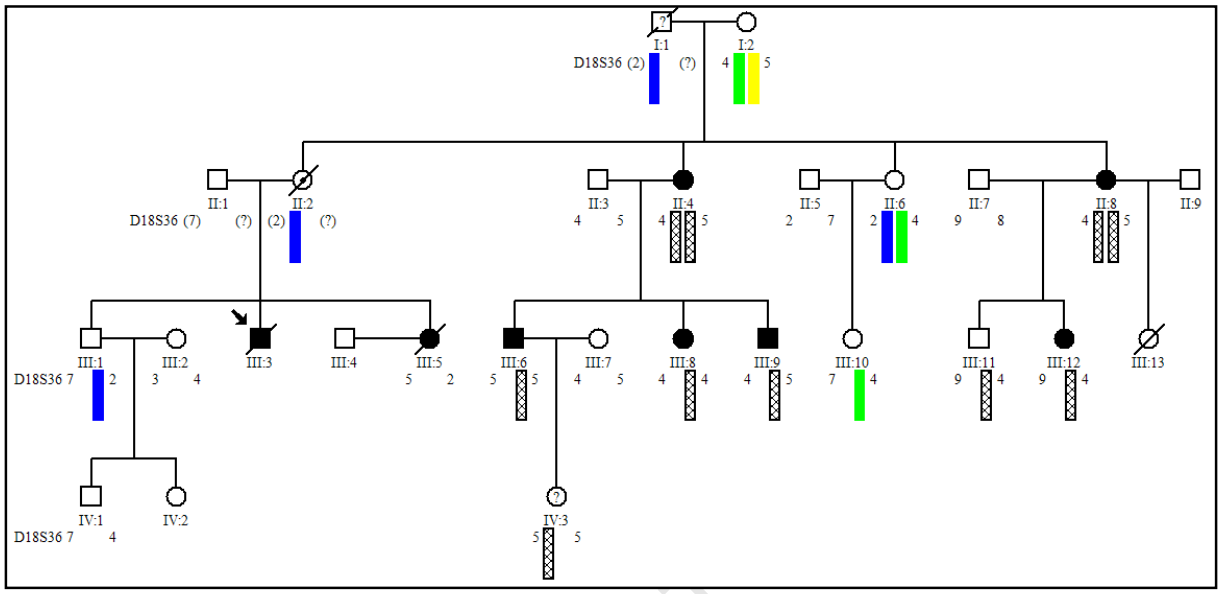


Figure 2-4: Pedigree of ACM 2 showing genotypes for the desmoglein-2 gene. Microsatellite marker names are shown at the left of each generation. Genotypes in brackets are inferred. The maternal allele is not indicated for family member III:5 as it is not possible to determine.

Table 2-5: Two-point lod scores obtained for the microsatellite marker flanking the desmoglein-2 gene.

		Lod score					
Marker	**Chromosomal position (in bp)**	**$\theta = 0$**	**$\theta = 0.1$**	**$\theta = 0.2$**	**$\theta = 0.3$**	**$\theta = 0.4$**	**$\theta = 0.5$**
D18S36	29,164,197 - 29,164,340	-2.52	-1.19	-0.61	-0.27	-0.09	0

It was not possible to determine which allele was inherited from which parent for family member II:4 and II:8. Despite this it was still possible to exclude desmoglein-2 as a candidate gene as no genotype is tracking with disease and a lod score less than -2 was observed at $\theta = 0$.

2.3.5 Desmocollin-2

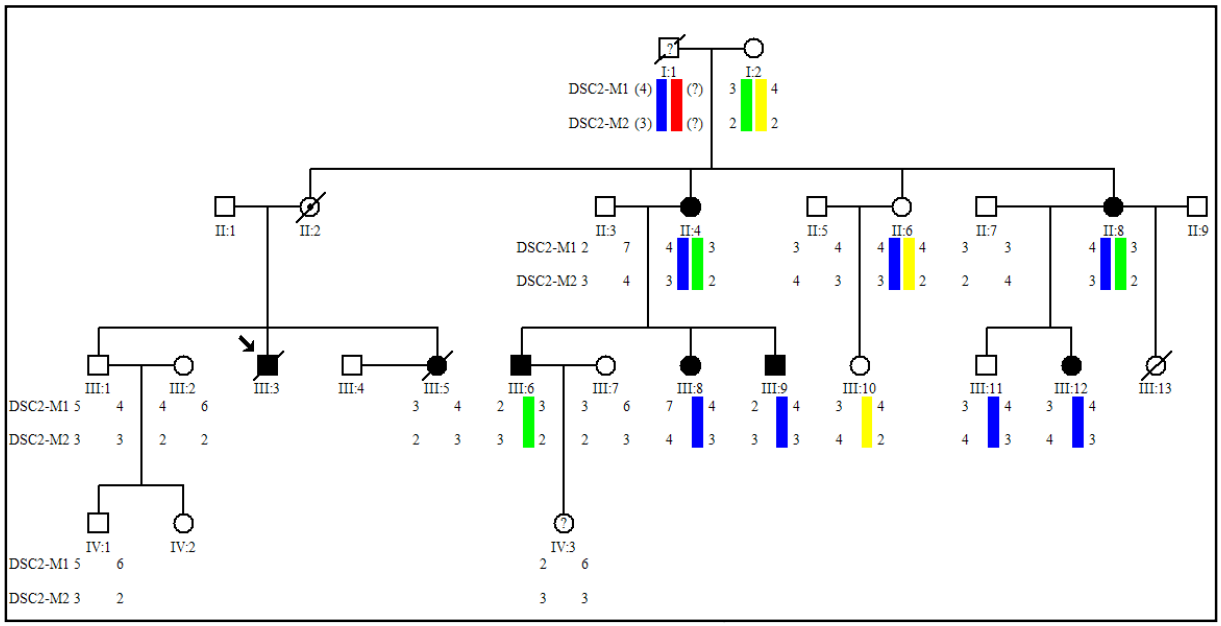


Figure 2-5: Pedigree of ACM 2 showing genotypes for the desmocollin-2 gene. Microsatellite marker names are shown at the left of each generation. Genotypes in brackets are inferred. Microsatellite marker AFM344ZD5 is not polymorphic in this family and therefore is not shown in the pedigree. The maternal allele is not indicated for family members III:1 and III:5 as it is not possible to determine.

Table 2-6: Two-point lod scores obtained for microsatellite markers flanking and within the desmocollin-2 gene.

		Lod score					
Marker	**Chromosomal position (in bp)**	**$\theta = 0$**	**$\theta = 0.1$**	**$\theta = 0.2$**	**$\theta = 0.3$**	**$\theta = 0.4$**	**$\theta = 0.5$**
DSC2-M1	28,717,808 - 28,718,093	-1.94	-0.74	-0.31	-0.10	-0.01	0
DSC2-M2	28,653,147 - 28,653,437 (in intron 12 of gene)	-3.38	-1.44	-0.72	-0.32	-0.10	0

Haplotype 43 (represented in blue) is prominent in this family but it is not tracking with disease. The lod score for DSC2-M2 (which lies within the gene) at $\theta = 0$ is less than -2 which excludes desmocollin-2 as a candidate gene.

2.3.6 Plakoglobin

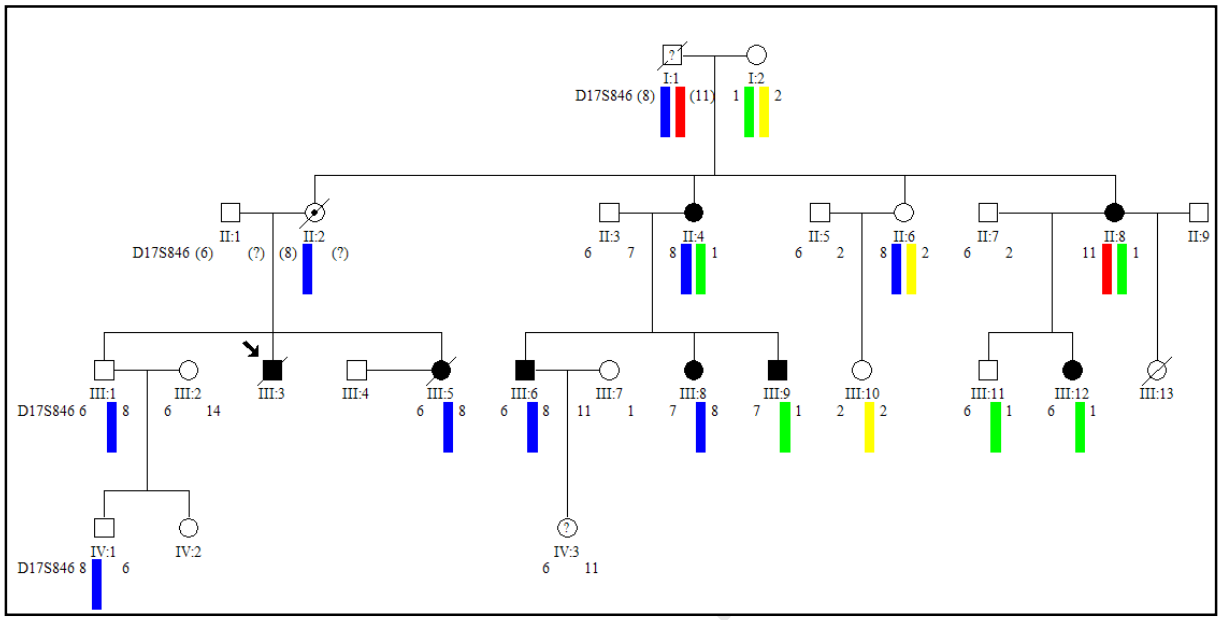


Figure 2-6: Pedigree of ACM 2 showing genotypes for the plakoglobin gene. Microsatellite marker names are shown at the left of each generation. Genotypes in brackets are inferred.

Table 2-7: Two-point lod scores obtained for the microsatellite marker flanking the plakoglobin gene.

		Lod score					
Marker	**Chromosomal position (in bp)**	**θ = 0**	**θ = 0.1**	**θ = 0.2**	**θ = 0.3**	**θ = 0.4**	**θ = 0.5**
D17S846	39,893,185 - 39,893,409	-4.26	-1.69	-0.69	-0.22	-0.02	0

Plakoglobin was excluded as a candidate gene as no genotype is tracking with disease and a lod score less than -2 was observed at θ = 0.

2.4 Discussion

It was considered prudent to perform exclusion mapping of the other known ARVC loci prior to positional candidate gene analysis of the ARVC-6 locus. The reason for this cautious approach was based on the less than definitive evidence for linkage of family ACM 2 to the ARVC-6 locus (Matolweni et al., 2006). Furthermore, erroneous linkage in ARVC has been known to occur. The ARVC-4 locus (2q32.1-q32.3) was identified based on the cumulative lod score of three small, unrelated ARVC families (Rampazzo et al., 1997). In one of these families a causal mutation was subsequently found in the plakophilin-2 gene on chromosome 12p11 (Gerull et al., 2004).

Exclusion mapping was performed in family ACM 2 at the ARVC-7 locus and for five ARVC disease genes. A total of 14 microsatellite markers were genotyped. Lod scores less than -2 were obtained for the ARVC-7 locus, desmoplakin, desmoglein-2, desmocollin-2 and plakoglobin. The lod score values obtained for plakophilin-2 do not support linkage of this locus with disease in this family. No haplotype or genotype (as is the case for desmoglein-2 and plakoglobin) segregated with ARVC in this family. The ARVC-7 locus and the five genes were therefore excluded from causing disease in this family.

Furthermore, an affected ACM 2 family member was included in a screen of South African ARVC patients for mutations in the five disease genes: plakophilin-2 (Watkins et al., 2009), desmoplakin (Fish, 2010), desmoglein-2, desmocollin-2 and plakoglobin (Mr Mzwandile Mbele, unpublished work). No disease-causing mutations were identified in this family member; this complementary approach supports the findings of the exclusion mapping of these five genes for this family.

Based on the findings of this study, the next step was positional candidate gene analysis of the ARVC-6 locus. It was decided to first prioritise the genes in this locus, for mutation screening, using bioinformatics tools.

3 PRIORITISATION OF CANDIDATE GENES

3.1 Introduction

Genome-wide linkage and association studies identify many candidate disease genes for further analysis (Tiffin et al., 2009). Subsequently, a number of bioinformatics tools have been developed to prioritise candidate disease genes for empirical research. A bioinformatics tool, GeneSeeker, selected *RMRP* as disease gene for skeletal dysplasia and it was confirmed to cause disease in patients (Thiel et al., 2005). Another bioinformatics tool, G2D, selected ten genes for asthma and atopy in a French Canadian population (Tremblay et al., 2008). A single nucleotide polymorphism (SNP) in one of the selected genes, *PTPRE*, was found to have a protective association with allergic asthma in this population. Bioinformatics tools have also been used in concert to prioritise candidate genes for type 2 diabetes (Tiffin et al., 2006), fetal alcohol syndrome (Lombard et al., 2007), osteoporosis (Huang et al., 2008) and metabolic syndrome (Tiffin et al., 2008). The advantage of using several bioinformatics tools together is to overcome the limitations inherent in each approach. It was therefore decided to use several existing bioinformatics tools to prioritise genes in the ARVC-6 locus for mutation screening.

3.2 Methods

3.2.1 Candidate gene list

Ensembl Gene IDs, for the genes located between microsatellite markers D10S1707 and D10S1477, were downloaded from the Ensembl database using BioMart (http://www.ensembl.org/biomart/index.html). The datasets used were Ensembl 41 and Homo sapiens genes (NCBI 36). A total of 26 Ensembl Gene IDs were retrieved. The physical map of the ARVC-6 locus is shown in Figure 3-1.

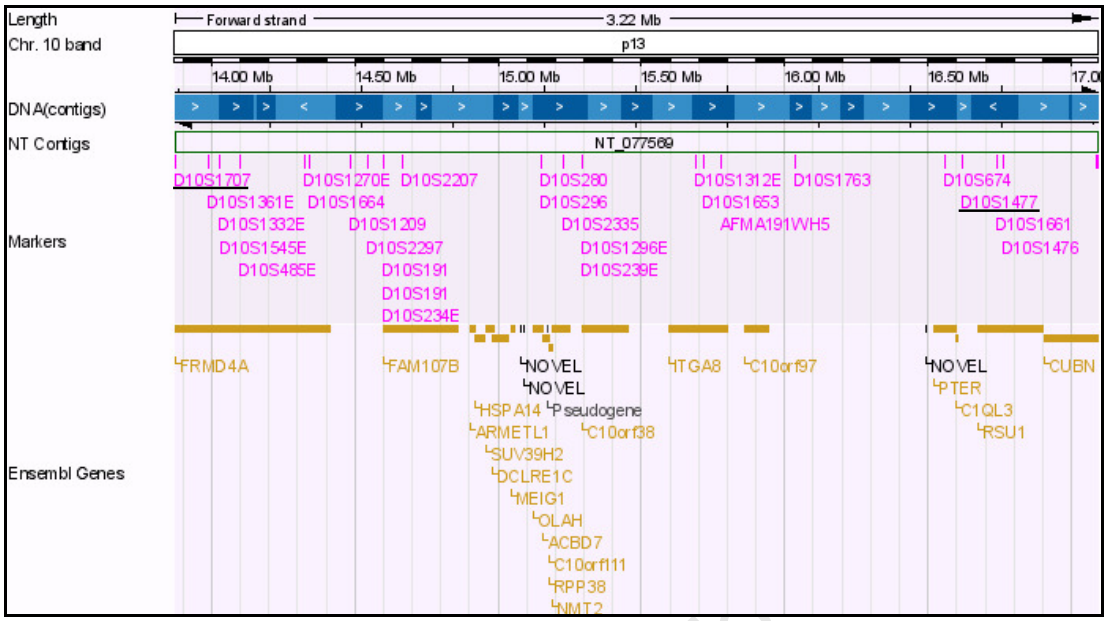


Figure 3-1: The physical map of the ARVC-6 locus, from microsatellite marker D10S1707 to D10S1477, captured from the Ensembl database.

3.2.2 eVOC method

This bioinformatics method uses eVOC anatomical ontology to combine text-mining of biomedical literature and data-mining of human gene expression data (Tiffin et al., 2005).

eVOC anatomy terms are extracted from PubMed abstracts containing the disease name of interest. The identified eVOC anatomy terms are ranked according to frequency of association with the disease name and frequency of annotation of genes with eVOC terms. *n* top-scoring eVOC anatomy terms are selected from the ranked list and compared with eVOC anatomy terms annotated to candidate disease genes. *m* mismatched terms are allowed (method described in full in Tiffin et al. 2005).

A total of 628 abstracts, containing the disease name "arrhythmogenic right ventricular cardiomyopathy" OR "arrhythmogenic right ventricular dysplasia", were downloaded from PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) on the 28th of November 2006.

Dragon Disease Explorer (DDE, http://research.i2r.a-star.edu.sg/DRAGON/DE/) was used to extract eVOC ontology terms in the abstracts. The frequency of association and frequency of annotation were calculated.

A python script **slurpee.py** (written by Alan Powell) was used to download all the information for the candidate genes from the Ensembl database and create a local database. The python scripts **real_disease_gene_finder.py** and **real_iterations.py** (written by Dr Nicki Tiffin) were run on the local database. Candidate genes that were annotated with at least three of the four top ranked eVOC anatomy terms (namely: "heart", "cardiovascular|vascular", "artery", "blood|peripheral blood") were selected as the most likely disease gene candidates (recommended by Tiffin et al. 2005).

3.2.3 GeneSeeker

GeneSeeker is a web-based tool (http://www.cmbi.ru.nl/GeneSeeker/) that gathers and combines positional data and expression/phenotypic data from web-based databases (van Driel et al., 2003; van Driel et al., 2005). The user specifies the "genetic location" and "expression location" for the disease of interest. GeneSeeker searches human and mouse genetic localisation, gene expression and phenotypic databases as listed in Table 3-1 (van Driel et al., 2003).

The results are given in four tables:

1. A list of human genes in the correct genetic region and matching the specified expression profile

2. A list of mouse genes matching the syntenic region(s) as well as the expression profile, but with no matching human gene name

3. A list of mouse genes found in the syntenic region, for which the homologous human gene is found to map outside the critical interval

4. A list of all the remaining human genes that are present in the genetic interval, but which do not match the expression profile (van Driel et al., 2003)

Table 3-1: Databases accessed by GeneSeeker

Localisation databases (human)	Localisation databases (mouse)	Expression/phenotype databases
OXFORD (srs.bioasp.nl:4080)	MGD (www.informatics.jax.org)	PubMed (www.ncbi.nlm.nih.gov/pubmed)
MIMMAP (srs.bioasp.nl:4080)		OMIM (srs.bioasp.nl:4080)
GDB (www.gdb.org)		UniProt (Swiss-Prot, TrEMBL, etc.) (srs.bioasp.nl:4080)
		GXD (www.informatics.jax.org)
		MLC (www.informatics.jax.org)
		TBASE (www.informatics.jax.org)

(van Driel et al., 2005)

The Ensembl database was used to determine the cytogenetic band on which the microsatellite markers reside. Both microsatellite markers were found on cytogenetic band 10p13. BioMart was used to check that all the genes (situated between the two microsatellite markers) are on 10p13. Various expression locations were entered i.e. heart, heart & ventricles, heart & right ventricle, heart & ventricle, cardiac. The python script **hugo_to_ensembl_id.py** (written by Dr Nicki Tiffin) was used to convert HUGO IDs (also known as HGNC symbols) to Ensembl Gene IDs.

3.2.4 Disease Gene Prediction (DGP)

Disease Gene Prediction (DGP, http://maine.ebi.ac.uk:8000/services/dgp) is a relational database that contains a probability score, of being involved in disease, for each gene present in Ensembl version 15.33.1. Probability scores were calculated based on the fact that proteins involved in hereditary diseases tend to be long, conserved, phylogenetically extended and without close paralogues (Lopez-Bigas and Ouzounis, 2004).

The website only allows submission of one Ensembl Gene ID at a time. A python script **screenscraper_dgp.py** (written by Dr Nicki Tiffin) was used to submit a list of Ensembl Gene IDs to the website. All 26 Ensembl Gene IDs were submitted to the website.

3.2.5 PROSPECTR

PROSPECTR (**PR**i**O**rization by **S**equence & **P**hylog**E**netic features of **C**andida**T**e **R**egions) is a web-based prioritisation method available at http://www.genetics.med.ed.ac.uk/prospectr/. It assigns a score (between 0 and 1) to each gene present in Ensembl version 27. The higher the score the higher the likelihood of involvement in disease (Adie et al., 2005). Scores were calculated based on eleven sequence features found to differ significantly between Ensembl genes found in Online Mendelian Inheritance in Man (OMIM) and those not in OMIM. These sequence features include signal peptide, mouse homolog % identity, length of 3' UTR, number of exons, rat homolog % identity, worm homolog % identity, GC, CDS length, gene length, mouse homolog Ka, paralog % identity (Adie et al., 2005).

The "score a single gene" option was chosen over the "score genomic region" as PROSPECTR is not maintained as was therefore using an out of date version of the Ensembl database i.e. version 27.

3.2.6 SUSPECTS

SUSPECTS (http://www.genetics.med.ed.ac.uk/suspects/) is a web-based server that prioritises positional candidate genes by combining annotation and sequence-based approaches (Adie et al., 2006). Two inputs are needed: the positional candidate genes and the known disease genes for the disorder (referred to as the "match set"). Each positional candidate gene is scored on four lines of evidence. PROSPECTR scores the gene on eleven **sequence-based features**. The gene is then scored on how well its **Gene Ontology (GO) terms** (Ashburner et al., 2000), **InterPro domains** (Hunter et al., 2009) and **gene expression profile** correlates with the match set. Each score is weighted depending on the amount of information available for each line of evidence. The four scores are combined to produce a final score ranging from 0 to 100. Higher scores represent better candidates

and the top score of 100 indicates a perfect match between the candidate gene and all genes in the match set (Adie et al., 2006).

A list of Ensembl Gene IDs were submitted to SUSPECTS instead of ranking genes between markers as SUSPECTS is not maintained as was therefore using an out of date version of the Ensembl database i.e. version v28.3. Two different match sets were submitted.

Match set 1: The five known ARVC genes that encode desmosomal proteins (*DSP*, *PKP2*, *DSG2*, *JUP* and *DSC2*)
Match set 2: All known ARVC genes at time of analysis (*DSP*, *PKP2*, *DSG2*, *JUP*, *DSC2*, *RyR2* and *TGFβ3*)

3.2.7 G2D

G2D (genes to diseases) is a web-based tool (http://www.ogic.ca/projects/g2d_2/) that prioritises genes in a region of interest using the disease phenotype, known genes or a second genomic region linked to the disease (Perez-Iratxeta et al., 2007). In this study the candidate genes were prioritised using the disease phenotype and known genes.

Disease phenotype:
The input for this approach is the region of interest and the OMIM identifier for the disease (Perez-Iratxeta et al., 2007). G2D prioritises the genes in the region of interest according to the description of the phenotype and its pre-computed associations to gene features as extracted from the literature and the Entrez Gene database.

Known genes:
This approach requires the region of interest and the Entrez Gene ID of one or several human or mouse genes that are already known or are suspected to be involved in the disease (Perez-Iratxeta et al., 2007). G2D prioritises the genes in the region of interest according to their similarity to the known gene(s) as given by their GO annotations and high sequence homology.

At time of analysis, the G2D website was being upgraded therefore the corresponding author (Carolina Perez-Iratxeta) was contacted for a list of GO IDs and scores associated with OMIM identifier ARVD-6 (604401) and the five known ARVC genes that encode desmosomal proteins (*DSP*, *PKP2*, *DSG2*, *JUP*, *DSC2*). The python script **g2d_go_analysis.py** (written by Dr Nicki Tiffin) used the supplied list of GO IDs and scores to score the candidate genes. The final score for each candidate gene is the maximum mapped GO score.

3.2.8 Endeavour

Endeavour is freely available software (http://www.esat.kuleuven.be/endeavour) that prioritises candidate "test" genes based on their similarity to known "training" genes (Aerts et al., 2006). Known genes are used to gather information about a disease or biological pathway from the following data sources:

A: literature (abstracts in EntrezGene)

B: functional annotation (Gene Ontology)

C: microarray expression (Atlas gene expression)

D: EST expression (EST data from Ensembl)

E: protein domains (InterPro)

F: protein-protein interactions (Biomolecular Interaction Network Database (BIND))

G: pathway membership (Kyoto Encyclopedia of Genes and Genomes (KEGG))

H: *cis*-regulatory modules (TOUCAN)

I: transcriptional motifs (TRANSFAC)

J: sequence similarity (BLAST)

K: additional data sources can be added

Candidate genes are then ranked based on their similarity with the training properties obtained from the data sources. This results in one prioritised list for each data source. The ranking from each data source is fused into a single ranking and provides an overall prioritisation for each candidate gene.

Ensembl Gene IDs for all known ARVC genes at time of analysis (*DSP*, *PKP2*, *DSG2*, *JUP*, *DSC2*, *RyR2* and *TGFβ3*) and the 26 candidate genes were loaded on Endeavour.

3.2.9 Tissue expression

Tissue expression data, from the GeneCards® database (http://www.genecards.org), was used to further prioritise the positional candidate genes. The tissue expression data is derived from three sources (GeneNote - expression arrays (Shmueli et al., 2003), UniGene - electronic Northern (Wheeler et al., 2003) and CGAP: SAGE (Lal et al., 1999)) and is represented in bar graphs. The genes received a score out of three, indicating the number of sources that showed normalised intensity/counts greater than 10 in the heart.

3.3 Results

3.3.1 eVOC method

Seven genes were selected by this method namely: *NMT2*, *FAM107B*, *C10orf97*, *HSPA14*, *SUV39H2*, *DCLRE1C* and *RPP38*.

3.3.2 GeneSeeker

NMT2 was the only gene selected by this method.

3.3.3 DGP

There was probability information for 15 of the 26 Ensembl Gene IDs submitted. The results are given in Table 3-2 below. The genes with a probability score greater than 0.5 (highlighted in yellow) were selected as potential disease-causing genes.

Table 3-2: Probability scores obtained for 15 candidate genes

Ensembl Gene ID	HGNC Symbol	Probability
ENSG00000065809	*FAM107B*	0.7243
ENSG00000185267	*ARMETL1*	0.474215
ENSG00000152455	*SUV39H2*	0.592058
ENSG00000152457	*DCLRE1C*	0.700131
ENSG00000152463	*OLAH*	0.385372
ENSG00000176244	*ACBD7*	0.392126
ENSG00000176236	*C10orf111*	0.307754
ENSG00000152464	*RPP38*	0.415991
ENSG00000152465	*NMT2*	0.354711
ENSG00000148468	*C10orf38*	0.18093
ENSG00000077943	*ITGA8*	0.669706
ENSG00000148481	*C10orf97*	0.489021
ENSG00000165983	*PTER*	0.53326
ENSG00000165985	*C1QL3*	0.599321
ENSG00000151474	*FRMD4A*	0.366396

3.3.4 PROSPECTR

There were scores for 14 of the 26 Ensembl Gene IDs submitted. The results are given in Table 3-3 below. The genes with a score greater than 0.5 (highlighted in yellow) were selected as potential disease-causing genes.

Table 3-3: Scores obtained for 14 candidate genes

Ensembl Gene ID	HGNC Symbol	Score
ENSG00000065809	*FAM107B*	0.42
ENSG00000187522	*HSPA14*	0.47
ENSG00000152455	*SUV39H2*	0.43
ENSG00000152457	*DCLRE1C*	0.52
ENSG00000152463	*OLAH*	0.33
ENSG00000176244	*ACBD7*	0.54
ENSG00000176236	*C10orf111*	0.15
ENSG00000152464	*RPP38*	0.43
ENSG00000152465	*NMT2*	0.52
ENSG00000148468	*C10orf38*	0.36
ENSG00000077943	*ITGA8*	0.60
ENSG00000148481	*C10orf97*	0.44
ENSG00000165983	*PTER*	0.52
ENSG00000151474	*FRMD4A*	0.54

3.3.5 SUSPECTS

Eight genes were not found in the hsapiens database and therefore could not be scored. Minor differences in rank order were observed when match set 1 was used compared to match set 2. The top 20% of genes scored were selected and are the same regardless of the match set used namely: *ITGA8*, *C1QL3*, *PTER* and *C10orf97*. Only the score and order of these genes differed.

3.3.6 G2D

The results of prioritisation using the disease phenotype and known genes are given in Table 3-4 and Table 3-5 respectively. All scored genes (highlighted in yellow) were selected as potential disease-causing genes.

Table 3-4: Scores obtained using the disease phenotype as input

Ensembl Gene ID	HGNC Symbol	Score
ENSG00000176244	*ACBD7*	0
ENSG00000165983	*PTER*	3.19e-006
ENSG00000204942	-	0
ENSG00000197889	*MEIG1*	1.76e-005
ENSG00000065809	*FAM107B*	0
ENSG00000201083	-	0
ENSG00000187522	*HSPA14*	2.3e-005
ENSG00000185221	-	0
ENSG00000148481	*C10orf97*	4.19e-006
ENSG00000148468	*C10orf38*	0
ENSG00000209435	-	0
ENSG00000152455	*SUV39H2*	1.76e-005
ENSG00000165985	*C1QL3*	1.25e-005
ENSG00000185267	*ARMETL1*	0
ENSG00000182531	-	0
ENSG00000201766	-	0
ENSG00000198508	-	0
ENSG00000151474	*FRMD4A*	4.27e-005
ENSG00000176236	*C10orf111*	4.61e-006
ENSG00000077943	*ITGA8*	1.98e-005
ENSG00000152465	*NMT2*	1.25e-005
ENSG00000201260	-	0
ENSG00000209429	-	0
ENSG00000152463	*OLAH*	0
ENSG00000152464	*RPP38*	1.76e-005
ENSG00000152457	*DCLRE1C*	1.76e-005

Table 3-5: Scores obtained using known genes as input

Ensembl Gene ID	HGNC Symbol	Score
ENSG00000176244	*ACBD7*	0
ENSG00000165983	*PTER*	2.15233
ENSG00000204942	-	0
ENSG00000197889	*MEIG1*	2.13895
ENSG00000065809	*FAM107B*	0
ENSG00000201083	-	0
ENSG00000187522	*HSPA14*	0
ENSG00000185221	-	0
ENSG00000148481	*C10orf97*	3.88469
ENSG00000148468	*C10orf38*	0
ENSG00000209435	-	0
ENSG00000152455	*SUV39H2*	2.13895
ENSG00000165985	*C1QL3*	2.13895
ENSG00000185267	*ARMETL1*	0
ENSG00000182531	-	0
ENSG00000201766	-	0
ENSG00000198508	-	0
ENSG00000151474	*FRMD4A*	4.76297
ENSG00000176236	*C10orf111*	1.52036
ENSG00000077943	*ITGA8*	5.16595
ENSG00000152465	*NMT2*	1.67061
ENSG00000201260	-	0
ENSG00000209429	-	0
ENSG00000152463	*OLAH*	0
ENSG00000152464	*RPP38*	2.13895
ENSG00000152457	*DCLRE1C*	1.73049

3.3.7 Endeavour

The top 25% of genes scored were selected namely: *DCLRE1C*, *ITGA8*, *PTER*, *C1QL3*, *C10orf97*, *FRMD4A* and *OLAH*.

3.3.8 Combined results for prioritisation methods

The results for each prioritisation method is summarised in Table 3-6. A gene scored 1 if it was prioritised by the method, 0 if it was not prioritised by the method and 0.5 if the method could not prioritise the gene due to lack of information. The scores were totalled and the candidate genes ranked according to total score.

Table 3-6: Combined prioritisation of candidate genes

Ensembl Gene ID	HGNC Symbol	eVOC method	GeneSeeker	DGP	PROSPECTR	SUSPECTS	G2D: disease phenotype	G2D: known genes	Endeavour	Total
ENSG00000152457	*DCLRE1C*	1	0	1	1	0	1	1	1	6
ENSG00000077943	*ITGA8*	0	0	1	1	1	1	1	1	6
ENSG00000165983	*PTER*	0	0	1	1	1	1	1	1	6
ENSG00000165985	*C1QL3*	0	0	1	0.5	1	1	1	1	5.5
ENSG00000152465	*NMT2*	1	1	0	1	0	1	1	0	5
ENSG00000148481	*C10orf97*	1	0	0	0	1	1	1	1	5
ENSG00000152455	*SUV39H2*	1	0	1	0	0	1	1	0	4
ENSG00000151474	*FRMD4A*	0	0	0	1	0	1	1	1	4
ENSG00000197889	*MEIG1*	0	0	0.5	0.5	0	1	1	0	3
ENSG00000152464	*RPP38*	1	0	0	0	0	1	1	0	3
ENSG00000187522	*HSPA14*	1	0	0.5	0	0	1	0	0	2.5
ENSG00000065809	*FAM107B*	1	0	1	0	0	0	0	0	2
ENSG00000176236	*C10orf111*	0	0	0	0	0	1	1	0	2
ENSG00000201766	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000182531	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000185221	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000209429	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000201260	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000209435	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000204942	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000201083	-	0	0	0.5	0.5	0.5	0	0	0	1.5
ENSG00000198508	-	0	0	0.5	0.5	0	0	0	0	1
ENSG00000152463	*OLAH*	0	0	0	0	0	0	0	1	1
ENSG00000176244	*ACBD7*	0	0	0	1	0	0	0	0	1
ENSG00000185267	*ARMETL1*	0	0	0	0.5	0	0	0	0	0.5
ENSG00000148468	*C10orf38*	0	0	0	0	0	0	0	0	0

3.3.9 Tissue expression

Only eight genes were scored based on tissue expression data and their score is given in brackets: *C10orf97* (3), *HSPA14* (3), *NMT2* (2), *FAM107B* (2), *PTER* (1), *RPP38* (1), *ARMETL1* (1) and *C10orf38* (1). These genes were moved to the top of the ranked list to give the final prioritised list shown in Table 3-7. Tissue expression of *ITGA8* and *FRMD4A* was not investigated and they were excluded from the final prioritised list as they were screened for mutations previously (Matolweni et al., 2006). The 16 genes without a score have normalised intensity/counts below 10 in the heart, are not expressed in the heart or do not have tissue expression data.

Table 3-7: Final prioritisation of candidate genes

Ensembl Gene ID	HGNC Symbol
ENSG00000148481	*C10orf97*
ENSG00000187522	*HSPA14*
ENSG00000152465	*NMT2*
ENSG00000065809	*FAM107B*
ENSG00000165983	*PTER*
ENSG00000152464	*RPP38*
ENSG00000185267	*ARMETL1*
ENSG00000148468	*C10orf38*
ENSG00000152457	*DCLRE1C*
ENSG00000165985	*C1QL3*
ENSG00000152455	*SUV39H2*
ENSG00000197889	*MEIG1*
ENSG00000176236	*C10orf111*
ENSG00000201766	-
ENSG00000182531	-
ENSG00000185221	-
ENSG00000209429	-
ENSG00000201260	-
ENSG00000209435	-
ENSG00000204942	-
ENSG00000201083	-
ENSG00000198508	-
ENSG00000152463	*OLAH*
ENSG00000176244	*ACBD7*

Note: genes highlighted in red = high expression in the heart
genes highlighted in blue = low expression in the heart

3.4 Discussion

A total of 26 Ensembl Gene IDs were downloaded from the Ensembl database, between microsatellite markers D10S1707 and D10S1477, using BioMart. Several existing bioinformatics tools and tissue expression data were used to prioritise the positional candidate genes for mutation screening.

The reason for using several bioinformatics tools was to overcome the limitations inherent in each approach. The utility of using bioinformatics tools in concert was extensively examined by Tiffin and colleagues (Tiffin et al., 2006). Methods that rely on annotation and biomedical literature (e.g. the eVOC method, GeneSeeker and G2D) are potentially biased towards selecting better studied genes. DGP and PROSPECTR score genes according to sequence features found to differ significantly between "disease genes" and "non-disease genes". While this approach is useful for identifying novel disease genes, genes that have been classified as "non-disease genes" might be found to cause disease in the future. Some methods require input of known disease genes to prioritise candidate genes (e.g. SUSPECTS, G2D and Endeavour). This approach is logical but restricts results to what is already known about the disease.

Subsequent to this study, a list of positional candidate genes was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) (21 genes) and was compared to the list downloaded from the Ensembl database (26 Ensembl gene IDs). Seventeen Ensembl gene IDs had a HGNC symbol and were present in the NCBI list. The remaining nine Ensembl gene IDs did not have a HGNC symbol and none matched the 3 pseudogenes and hypothetical protein remaining in the NCBI list. These nine Ensembl gene IDs genes were therefore removed from the final prioritised list.

ITGA8 and *FRMD4A* were previously prioritised for mutation screening by Matolweni and colleagues because they were found to have cell-to-cell adhesion properties and five of the ARVC disease genes are involved in cell-to-cell adhesion (Matolweni et al., 2006). In this study, *ITGA8* and *FRMD4A* were prioritised second and eighth respectively in the ranked list (Table 3-6). No disease-causing mutations were however identified in these two genes (Matolweni et al., 2006) and they were therefore not included in the final prioritised list (Table 3-7). This left a total of fifteen genes prioritised for mutation screening.

4 MUTATION SCREENING

4.1 Introduction

The aim of this study was to screen the coding regions and intron-exon boundaries of 15 genes in order of the final prioritised list (Chapter 3).

4.2 Methods

4.2.1 Mutation Screening by Sequencing

Eight genes were screened for mutations by sequencing namely: *C10orf97*, *HSPA14*, *NMT2*, *FAM107B*, *MEIG1*, *C10orf111*, *OLAH* and *ACBD7*.

4.2.1.1 PCR

4.2.1.1.1 Primer Design

Primers were designed for all the exons of the eight genes. According to NCBI, *OLAH* has two isoforms. Primers were therefore designed to screen the exons of both isoforms. An annotation program ANNOTV9 (designed by Dr George Rebello) was used to designate the exons and reported SNPs in the genome sequence. Primer 3 was used to design the primers. OligoAnalyzer 3.0 and BLAST (Basic Local Alignment Search Tool) were used to evaluate the primers for *C10orf97*.

OligoAnalyzer 3.0 (http://scitools.idtdna.com/scitools/Applications/OligoAnalyzer/Default.aspx) was used to identify possible:

1. Hairpin structures: annealing of the primer with itself

2. Homo-dimers: annealing of primers of the same type

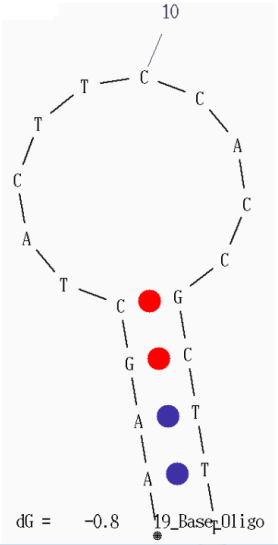
3. Hetero-dimers: annealing within primer pairs

OligoAnalyzer 3.0 produces a graphical representation and a delta g value for each possible secondary structure of a primer. An example of each secondary structure is given in Figure 4-1. If the delta g value was -9kcal/mole or more negative, the primer was redesigned (OligoAnalyzer 3.0 technical support, personal communication).

BLAST (http://blast.ncbi.nlm.nih.gov/) was used to search for the primer in the human genome. If the 3' end of the primer was found anywhere else in the genome, the primer was redesigned.

In order to save time, reverse e-PCR was used to evaluate the remaining primers instead of OligoAnalyzer 3.0 and BLAST. Reverse e-PCR was used to search for the primer pair in the human genome. If the primer pair was found anywhere else in the genome, the primers were redesigned. The sequence of these primers (Department of Molecular and Cell Biology, UCT or Whitehead Scientific) is given in Appendix B.

A



B

HOMO-DIMER ANALYSIS

Dimer Sequence

5'- AAGCTACTTCCACCGCTTT -3'

Maximum Delta G -37.37 kcal/mole

Delta G -6.68 kcal/mole
Base Pairs 4

```
5'  AAGCTACTTCCACCGCTTT
    ||||          ::::
3' TTTCGCCACCTTCATCGAA
```

C

HETERO-DIMER ANALYSIS

Primary Sequence

5'- AAGCTACTTCCACCGCTTT -3'

Secondary Sequence

5'- GTGTCTAGAAAGTCCAGCTTCA -3'

Maximum Delta G -37.37 kcal/mole

Delta G -8.28 kcal/mole
Base Pairs 5

```
5'   AAGCTACTTCCACCGCTTT
     |||||       :
3' ACTTCGACCTGAAAGATCTGTG
```


Figure 4-1: Secondary structure prediction by OligoAnalyzer 3.0. **A.** Hairpin structure **B.** Homo-dimer **C.** Hetero-dimer.

4.2.1.1.2 PCR Optimisation

PCR was optimised using the temperature gradient function on the Labnet MultiGene Thermal Cycler (Whitehead Scientific) (Appendix C). The PCR products were separated by gel electrophoresis through a 1% agarose gel and visualised using ethidium bromide (Appendix D). A volume of 5µl of PCR product was mixed with 3µl agarose loading dye (Appendix E) and electrophoresed alongside a DNA ladder (0.5µg 1 Kb Plus DNA Ladder (*Invitrogen*, Laboratory Specialist Services), 0.5µg DNA Molecular Weight Marker XIV (*Roche*) or 650ng 100bp DNA Ladder (*Promega*, Whitehead Scientific) (Appendix F)). The annealing temperature chosen for PCR of each exon is given in Appendix B.

4.2.1.1.3 Amplification of Patient DNA

Exons were amplified in two affected and two unaffected family members using the optimised PCR conditions (Appendix C). A negative control, where dH_2O replaced DNA, was included with PCRs set up to rule out contamination. PCR cycling was carried out on a Labnet MultiGene Thermal Cycler (Whitehead Scientific) at the cycling conditions given in Appendix C and the annealing temperatures in Appendix B.

To confirm amplification, PCR products were separated by gel electrophoresis through a 1% agarose gel and visualised using ethidium bromide (Appendix D). A volume of 5µl of PCR product was mixed with 3µl agarose loading dye (Appendix E) and electrophoresed alongside a DNA ladder (0.5µg 1 Kb Plus DNA Ladder (*Invitrogen*, Laboratory Specialist Services), 0.5µg DNA Molecular Weight Marker XIV (*Roche*) or 650ng 100bp DNA Ladder (*Promega*, Whitehead Scientific) (Appendix F)).

4.2.1.2 Sequencing

Purification of PCR products was performed in 0.2ml tubes (Whitehead Scientific) in a total of 20µl containing 5 or 10µl of PCR product, 1 unit of Shrimp Alkaline Phosphatase (SAP) (*Promega*, Whitehead Scientific), 2 units of Exonuclease I (Exo I) (*New England Biolabs*, Laboratory Specialist Services) made up with dH_2O. Exo I removes the leftover primers and the SAP removes the dNTPs. The tubes were incubated in a Labnet

MultiGene Thermal Cycler (Whitehead Scientific) for 60 minutes at 37°C to degrade the remaining primers and dNTPs followed by 15 minutes at 75°C to inactivate the enzymes. Primers were diluted to 1.1ng/µl and sent with the purified PCR products to the Central Analytical Facility, Stellenbosch University for sequencing. Sequencing was performed in both directions to screen for mutations in the exons.

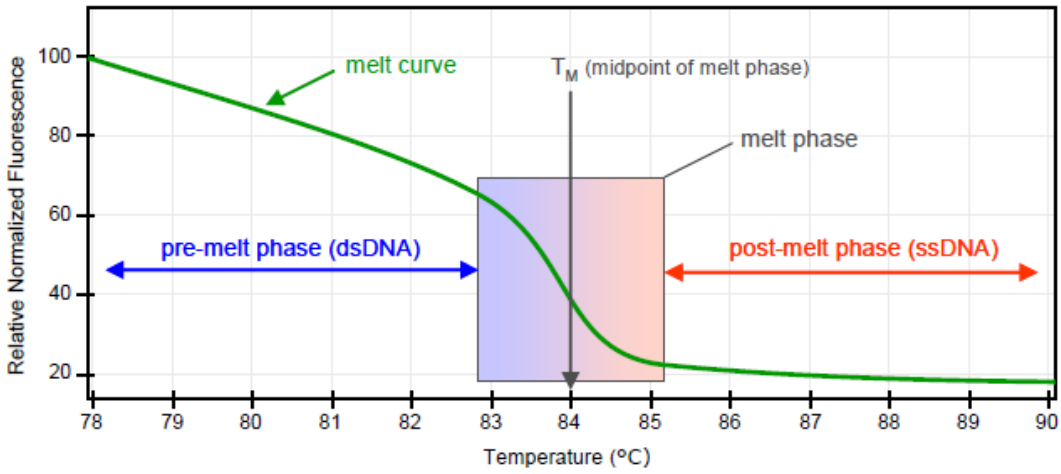
The electropherograms were analysed by eye to identify heterozygous variants. To identify homozygous variants, the patients' sequences were aligned to the wild type sequence. BioEdit version 7.0.0 (http://www.mbio.ncsu.edu/bioedit/bioedit.html) was used to align the sequences and the wild type sequence was obtained from NCBI.

4.2.2 Mutation Screening by High Resolution Melt (HRM)

High Resolution Melt (HRM), performed by the Rotor-Gene™ 6000 (*Corbett Research*, Celtic Molecular Diagnostics), was used to screen for mutations in *NMT2* exon 1. This exon is 71% GC-rich and was difficult to sequence.

4.2.2.1 Principle of HRM

HRM analysis characterises DNA sequences based on their dissociation (melt) profiles. The DNA sequence is amplified by PCR in the presence of a double stranded DNA (dsDNA) intercalating fluorescent dye (e.g. EvaGreen™). The dye intercalates with dsDNA and fluoresces brightly. The PCR products are subjected to a melt step and they transition from dsDNA to single stranded DNA (ssDNA) with increasing temperature. The dye is released and the fluorescence decreases. The Rotor-Gene™ 6000 collects the fluorescent signals. Figure 4-2 shows a typical HRM plot. The melt behaviour of a sequence depends on its composition and length (HRM Assay Design and Analysis booklet).



(HRM Assay Design and Analysis booklet)

Figure 4-2: A typical HRM plot.

4.2.2.2 Primer Design

Two extra primers were designed to divide the *NMT2* exon 1 amplicon into two overlapping amplicons (*NMT2* exon 1 amplicon 1 and 2). OligoAnalyzer 3.0 and BLAST were used to evaluate the primers. The sequence of primers NMT2Exon1F2 & R2 (Whitehead Scientific) are given in Appendix B.

4.2.2.3 PCR and HRM Optimisation

PCR and HRM were optimised for *NMT2* exon 1 amplicon 1 and 2 on the Rotor-Gene™ 6000 (*Corbett Research*, Celtic Molecular Diagnostics) using two control DNA samples for each amplicon (Appendix C). A suitable profile was not obtained for *NMT2* exon 1 amplicon 1 so this amplicon was screened for mutations by sequencing (Appendix C).

4.2.2.4 Mutation Screening of Patients

NMT2 exon 1 amplicon 2 was amplified in two affected and two unaffected family members using the optimised PCR and HRM conditions (Appendix C). A positive control (one of the control DNA samples used in optimisation) and a negative control (dH_2O replaced DNA to rule out contamination) were included.

4.2.3 Mutation Screening by Denaturing High Performance Liquid Chromatography (DHPLC)

Seven genes were screened for mutations by DHPLC namely: *PTER*, *RPP38*, *ARMETL1*, *C10orf38*, *DCLRE1C*, *C1QL3* and *SUV39H2*.

4.2.3.1 PCR

4.2.3.1.1 Primer Design

Primers were designed for all the exons of the seven genes. According to NCBI, *PTER* and *RPP38* both have two isoforms. Primers were therefore designed to screen the exons of both isoforms. ANNOTV9 was used to annotate the genome sequence with exons and SNPs, Primer 3 was used to design the primers and reverse e-PCR was used to evaluate the primers. The sequence of these primers (*Invitrogen*, Italy) is given in Appendix B.

4.2.3.1.2 PCR Optimisation

PCR was optimised using the temperature gradient function on the Mastercycler® pro (*Eppendorf*) (Appendix C). The PCR products were separated by gel electrophoresis through a 2% agarose gel and visualised using ethidium bromide (Appendix D). A volume of 5µl of PCR product was mixed with 1µl ReddyRun^{TM} gel loading buffer (*ABgene*) and electrophoresed alongside either 0.15µg Superladder-Low 100bp Ladder with ReddyRun^{TM} (*ABgene*) or 0.3µg 100bp DNA Ladder (*GENECRAFT*) (Appendix F). The annealing temperature chosen for PCR of each exon is given in Appendix B.

4.2.3.1.3 Amplification of Patient DNA

Exons were amplified in duplicate in a positive control (the control DNA sample used in optimisation), two affected and two unaffected family members using the optimised PCR conditions (Appendix C). A negative control, where dH_2O replaced DNA, was included with PCRs set up and run on the WAVE machine to rule out contamination. PCR cycling was carried out on the Mastercycler® pro (*Eppendorf*) at the cycling conditions given in Appendix C and the annealing temperatures in Appendix B.

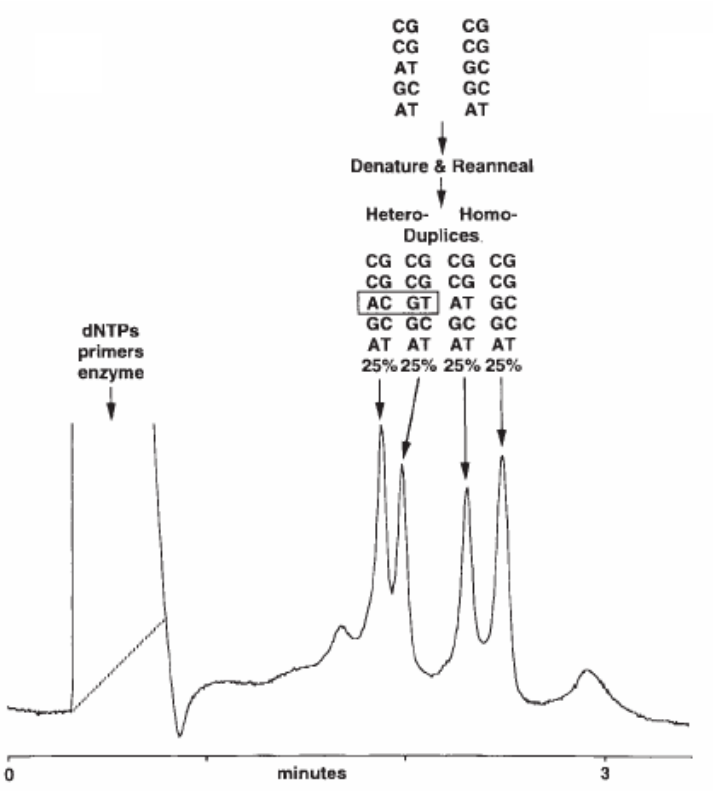
4.2.3.2 DHPLC

DHPLC, performed by the WAVE^{R} System (*Transgenomic*), was used to screen for mutations in the exons. This method was initially developed in 1995 by Oefner and Underhill (Oefner and Underhill, 1995).

4.2.3.2.1 Principle of DHPLC

DHPLC, under partial denaturing conditions, can detect single base substitutions, small insertions and deletions (Xiao and Oefner, 2001). Prior to DHPLC analysis, PCR products are denatured at 95^{o}C for 5 minutes and re-annealed by gradual cooling. If one of the chromosomes contains a sequence variant, the PCR products will re-anneal to form the original homoduplexes and the sense and anti-sense stands from different PCR products will form heteroduplexes (Xiao and Oefner, 2001).

The PCR products are loaded on the WAVE machine which injects the sample onto the separation cartridge. The cartridge is filled with alkylated nonporous polystyrene-divinylbenzene (PS-DVB) beads which are electrostatically neutral and hydrophobic. To aid adsorption of nucleic acids to the beads, triethylammonium acetate (TEAA) is added. This reagent is positively charged and hydrophobic therefore it interacts with nucleic acids (which are negatively charged) and the hydrophobic beads. Increasing concentrations of acetronitrile flows across the cartridge. The hydrophobic interaction between the cartridge and DNA/TEAA is broken and the PCR product elutes (WAVE^{R} manual). Under partial denaturing conditions, heteroduplexes denature more extensively at the analysis temperature and therefore elute earlier than the homoduplexes that undergo less pronounced denaturation. The ideal chromatographic profile is illustrated in Figure 4-3. The order of elution of the four species is primarily determined by neighbouring stacking interactions (Xiao and Oefner, 2001).



(Xiao and Oefner, 2001)

Figure 4-3: An ideal chromatographic profile for a mismatch.

4.2.3.2.2 Design of WAVE methods

A WAVE method consists of an analysis temperature and DHPLC conditions. WAVEMAKER™ software was used to identify the melting domains in an amplicon and a WAVE method was designed for each domain. The analysis temperature for a domain was chosen where the helical fraction (i.e., double strandedness) of the domain is between 75 - 100%. The application type "mutation detection" was selected.

The WAVE methods were tested on the control sample that amplified the best during PCR optimisation. A volume of 10µl of PCR product was injected onto the cartridge for each WAVE method. The temperatures were adjusted or fewer temperatures were used based on the test run. Suitable chromatographic profiles were not obtained for *RPP38* exon 2 so this exon was screened for mutations by sequencing.

4.2.3.2.3 Mutation Screening of Patients

Heteroduplex formation was carried out on the Mastercycler® pro (*Eppendorf*). The samples were denatured at 95°C for 5 minutes and then the temperature was decreased by 1°C every minute from 95°C to 60°C. The samples then stood for 10 minutes at room temperature.

4.2.3.3 Sequencing

All amplicons with aberrant chromatographic profiles were sequenced in both directions (Appendix C).

4.3 Results

A total of 36 sequence variants (SVs) were identified in the 15 genes. Twenty-nine SVs have been reported as SNPs on the NCBI SNP database and seven SVs are novel. Only one SV, c.1641A>G in *DCLRE1C*, was found in the two affected family members and was not present in the two unaffected family members screened. This single base substitution results in a synonymous amino acid change. A summary of the SVs is given in Table 4-1.

Table 4-1: A summary of the sequence variants identified

HGNC Symbol	Accession number	Variant Nomenclature	SNP accession number	Classification
C10orf97	NM_024948.2	c.-5T>C	rs2297882	Noncoding
C10orf97	NM_024948.2	c.731-23G>T	-	Noncoding
C10orf97	NM_024948.2	c.956-90G>A	-	Noncoding
HSPA14	NM_016299.2	c.-8G>T	-	Noncoding
HSPA14	NM_016299.2	c.-7C>T	-	Noncoding
NMT2	NM_004808.2	c.1170+68T>G	rs10906832	Noncoding
NMT2	NM_004808.2	c.1171-12T>C	rs7072269	Noncoding
NMT2	NM_004808.2	c.1339-69A>C	rs7918775	Noncoding
FAM107B	NM_031453.2	c.*92T>C	rs3780915	Noncoding
PTER	NM_001001484.1/ NM_030664.3	c.-222C>G/ c.-167C>G	rs7922050	Noncoding
PTER	NM_001001484.1/ NM_030664.3	c.-182-105C>G/ c.-128+205C>G	rs7894216	Noncoding
PTER	NM_001001484.1/ NM_030664.3	c.-182-100T>C/ c.-128+210T>C	rs7910712	Noncoding
PTER	NM_001001484.1/ NM_030664.3	c.-48-17127C>T c.-49C>T	rs6602121	Noncoding
PTER	NM_001001484.1/ NM_030664.3	c.432+25G>A	rs2275731	Noncoding
RPP38	NM_183005.3/ NM_006414.3	c.-430G>A/ c.-11+57G>A	-	Noncoding
RPP38	NM_183005.3/ NM_006414.3	c.-248C>T/ c.-11+239C>T	rs7900561	Noncoding
ARMETL1	NM_001029954.2	c.306A>C	rs61843027	Synonymous (P>P)
C10orf38	NM_001010924.1	c.418+19delT	-	Noncoding
C10orf38	NM_001010924.1	c.1393C>T	rs3814165	Missense (P>S)
DCLRE1C	**NM_001033858.1**	**c.1641A>G**	**-**	**Synonymous (L>L)**

MEIG1	NM_001080836.2	c.-29-307A>G	rs6602774	Noncoding
MEIG1	NM_001080836.2	c.-29-216T>C	rs4750567	Noncoding
MEIG1	NM_001080836.2	c.-29-132T>C	rs4237442	Noncoding
MEIG1	NM_001080836.2	c.26A>C	rs4750568	Missense (K>T)
MEIG1	NM_001080836.2	c.139-94C>G	rs1935400	Noncoding
MEIG1	NM_001080836.2	c.139-72A>G	rs1935399	Noncoding
MEIG1	NM_001080836.2	c.*58C>T	rs4418687	Noncoding
MEIG1	NM_001080836.2	c.*62A>G	rs4295952	Noncoding
MEIG1	NM_001080836.2	c.*204T>C	rs5019235	Noncoding
MEIG1	NT_008705.16	g.14954935A>G	rs11259410	Noncoding
MEIG1	NT_008705.16	g.14954964A>C	rs11259411	Noncoding
C10orf111	NM_153244.1	c.209G>A	rs7896053	Missense (R>K)
C10orf111	NM_153244.1	c.*92A>T	rs6602812	Noncoding
OLAH	NM_018324.2/ NM_001039702.2	c.501A>G/ c.342A>G	rs10906818	Synonymous (L>L)
ACBD7	NM_001039844.2	c.12+41G>A	rs12261006	Noncoding
ACBD7	NM_001039844.2	c.130+58G>A	rs4750594	Noncoding

Note: The SV found exclusively in the two affected family members screened is written in bold.

4.4 Discussion

A comprehensive mutation screening study was conducted for 15 genes in the ARVC-6 region using different methods. The top four prioritised genes (*C10orf97*, *HSPA14*, *NMT2* and *FAM107B*) were considered important candidate genes due to their high expression in the heart. These genes were therefore screened using a definitive method of mutation detection i.e. sequencing. *NMT2* exon 1 was problematic to sequence due to its high GC content. Two extra primers were designed to divide the *NMT2* exon 1 amplicon into two overlapping amplicons (*NMT2* exon 1 amplicon 1 and 2). *NMT2* exon 1 amplicon 1 was screened by sequencing and *NMT2* exon 1 amplicon 2 was screened by HRM. The next seven genes on the prioritised list (*PTER*, *RPP38*, *ARMETL1*, *C10orf38*, *DCLRE1C*, *C1QL3* and *SUV39H2*) were screened for mutations by DHPLC which is cost-effective,

rapid and nearly 100% of sequence variations can be detected (Oefner and Underhill, 1999; Xiao and Oefner, 2001). The remaining four genes (*MEIG1*, *C10orf111*, *OLAH* and *ACBD7*) had few, small exons and were therefore screened for mutations by sequencing.

A total of 36 SVs were identified. Twenty-nine SVs have been reported as SNPs on the NCBI SNP database and seven SVs are novel. Only one SV, c.1641A>G in *DCLRE1C*, was found in the two affected family members and was not present in the two unaffected family members screened. This substitution results in a new codon but it specifies the same amino acid (i.e. synonymous substitution).

Some synonymous substitutions have been found to affect pre-mRNA splicing (Cartegni et al., 2002). Incorrect splicing creates aberrant mRNAs that are either unstable or code for defective or deleterious protein isoforms. In ARVC, a homozygous synonymous substitution was identified in *PKP2* that causes cryptic splicing (Awad et al., 2006). In this study, it is therefore important to screen the synonymous substitution found in *DCLRE1C* in the rest of family ACM 2. If the SV is found to track with disease in the family then its effect on pre-mRNA splicing will need to be experimentally determined.

While no overt disease-causing mutations were identified in the coding regions and intron-exon boundaries, mutations in the 5' and 3' UTRs cannot be ruled out. Beffagna and colleagues reported disease-causing mutations in the 5' and 3' UTRs of the *TGFβ3* gene (Beffagna et al., 2005). In a canine model of ARVC, dogs homozygous for an 8bp deletion in the 3' UTR of the striatin gene had a more severe form of disease (Meurs et al., 2010).

Two positional candidate genes, *PTPLA* (Li et al., 2000b) and Human *NAPOR* (*CUGBP2*) (Li et al., 2001), were screened in the North American family in which the ARVC-6 locus was originally mapped (Li et al., 2000a). No disease-causing mutations were identified. These two genes are not in the linkage region of family ACM 2 and therefore were not screened in this study.

The mutation screening methods employed in this study (i.e. sequencing, HRM and DHPLC) would not have identified copy-number variations (CNVs). It was therefore decided to screen for CNVs in this locus.

5 MUTATION SCREENING FOR COPY-NUMBER VARIANTS

5.1 Introduction

A copy-number variation (CNV) is a segment of DNA that represents an imbalance between two genomes from one species (Stankiewicz and Lupski, 2010). Deletions, duplications, triplications, insertions or translocations can result in CNVs. This structural variation has been implicated in Mendelian diseases (e.g. Duchenne muscular dystrophy), common complex traits (e.g. autism) and disease susceptibility (e.g. psoriasis).

It was decided to screen for CNVs in the ARVC-6 region using the Affymetrix® Genome-Wide Human SNP Array 6.0 (hereafter referred to as SNP Array 6.0). The SNP Array 6.0 contains 906,600 single nucleotide polymorphisms (SNPs) and 946,000 non-polymorphic probes for the detection of copy-number variation with a combined intermarker genome spacing of 680 bases (Affymetrix® Genome-Wide Human SNP Array 6.0 Data Sheet and Affymetrix® FAQs).

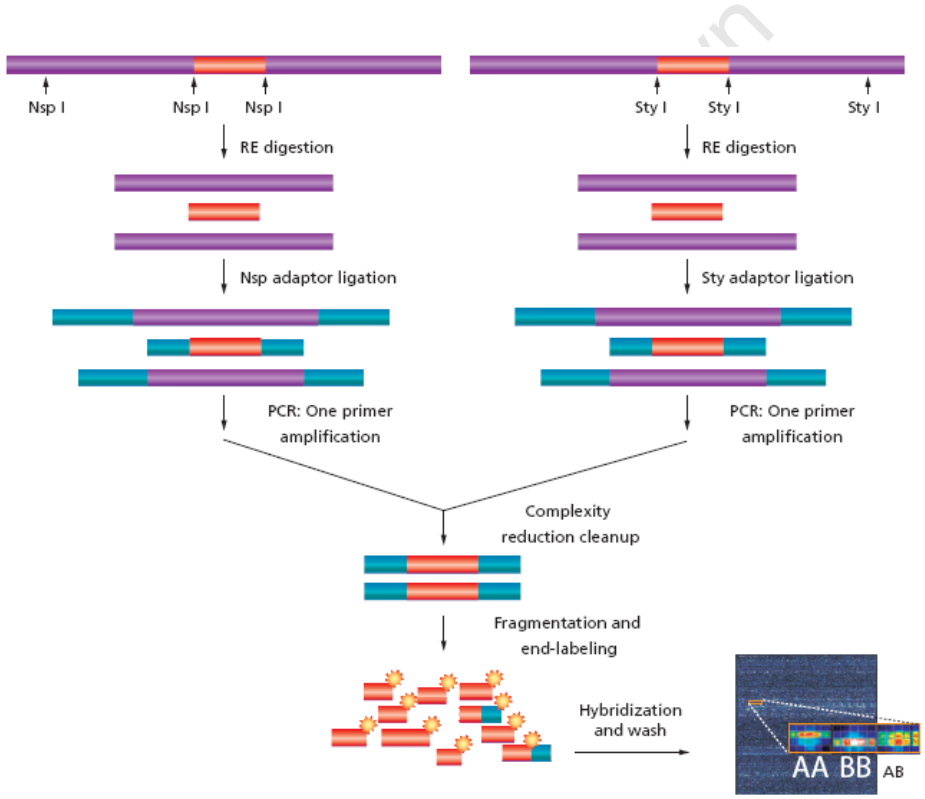
5.2 Method

5.2.1 SNP Array 6.0

A DNA sample from one affected family member (III:9 sample C) was sent to the Centre for Proteomic and Genomic Research (CPGR), University of Cape Town for genotyping with the SNP Array 6.0. Family member III:9 was one of the affected family members screened for mutations in Chapter 4. DNA sample C had sufficient quantity and quality of DNA. This specific sample was also genotyped for exclusion mapping and sample mix up can therefore be ruled out.

The CPGR determined the concentration and assessed the integrity of the DNA sample. The DNA was diluted to a working concentration of 50ng/µl and the Affymetrix® Genome-Wide Human SNP Assay 5.0/6.0 protocol was followed. A total of 500ng DNA was digested with Nsp I and Sty I restriction enzymes and ligated to adaptors. Adaptor-

ligated DNA fragments 200 to 1,100bp in size were preferentially amplified using optimised PCR conditions and a primer that recognises the adaptor sequence. The PCR products, for each restriction enzyme digest, were combined and purified using polystyrene beads. The amplified DNA was fragmented, labelled with biotin and hybridised to the SNP Array 6.0 (Affymetrix® Genome-Wide Human SNP Array 6.0 Data Sheet). The array was rinsed to remove DNA that did not bind. A fluorescent stain was added and the fluorescent molecules adhered to the biotin (Structure & Function of GeneChip® Microarrays). The results were generated by laser-guided optical scanning of the array. Figure 5-1 gives an overview of the SNP Assay 5.0/6.0.



(Affymetrix® Genome-Wide Human SNP Array 6.0 Data Sheet)

Figure 5-1: Workflow of the Affymetrix® Genome-Wide Human SNP Assay 5.0/6.0

5.2.2 Quality Control

Quality control for the genotyping and copy-number analysis was performed by the CPGR using the Birdseed version 2 and Canary version 2 algorithms respectively on the Affymetrix Genotyping Console® 4.0.

5.2.3 Copy-Number Analysis

Copy-number analysis was performed by the CPGR using the Canary version 2 algorithm on the Affymetrix Genotyping Console® 4.0 and a reference model file provided by Affymetrix (GenomeWideSNP_6.hapmap270.na30.r1.a5.ref). The following segment reporting parameters were used: minimum number of markers per segments: 5, minimum genomic size of segment: 0.1kb and include segments that overlap with known CNV: 100%.

Analysis was also performed changing the following Canary parameters sequentially:

1. Transition Decay from 1000Mb to 10Mb
2. CN Minimum Number Marker Threshold from 1 to 0
3. Regional GC correction to no regional GC correction

The default settings of these three parameters model the data to obtain the best fit regardless of experimental error and array inconsistencies. The three parameters were changed sequentially to obtain the unmodelled data and possibly identify regions that were missed during the initial analysis.

5.3 Results

The sample had a contrast QC (quality control measure of genotyping analysis) of 2.49 (Affymetrix recommended threshold should be greater than 1.7). The sample had a median absolute pairwise difference (MAPD) (quality control measure of copy-number analysis) of 0.28 (Affymetrix recommended threshold should be less than 0.35). The sample therefore passed both quality control measures.

No CNV was found in the ARVC-6 region despite changing various parameters i.e. transition decay, CN minimum number marker threshold and selecting no regional GC correction.

5.4 Discussion

Copy-number analysis, of one affected family member, did not identify any CNVs in the ARVC-6 region. Copy-number analysis was performed for the whole genome which is an unbiased approach to data analysis. The results of only the ARVC-6 region were reported as this locus shows linkage in family ACM 2. The Canary parameters were set to identify known, partially known and novel CNVs.

A similar approach as this study was used to successfully identify the genetic cause of an autoinflammatory disease in a single patient (Reddy et al., 2009). Genomewide copy-number analysis was performed using a SNP Array 6.0, the Affymetrix Genotyping Console® 2.0 and a reference model (GenomeWideSNP_6.hapmap270.422 data set). A 175kb homozygous deletion at chromosome 2q13 was identified, in the patient, which encompasses six interleukin-1 family members. The consanguineous parents were heterozygous for this deletion and were healthy.

Since no CNVs were identified in the ARVC-6 region, two different approaches were considered as the next experiment: whole-genome genetic linkage analysis or sequencing the ARVC-6 region using next-generation sequencing (NGS).

NGS are newer methods of sequencing compared to the automated Sanger method ("first-generation") (Metzker, 2010). These newer methods are able to produce large volumes of sequence data cheaper than the Sanger method. NGS is commercially available from Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences. Each company employs different strategies for NGS but in general the steps involved are template preparation, sequencing & imaging and data analysis. Sequencing the whole ARVC-6 region was considered as this would identify all variation in the region.

In the end, whole-genome genetic linkage analysis was chosen as the next experiment. This approach was considered the sensible option as linkage to only the known ARVC loci was performed in family ACM 2 and the lod score obtained for the ARVC-6 locus (2.93) was below the classical threshold of significance (Z = 3.0).

6 WHOLE-GENOME SNP LINKAGE ANALYSIS

6.1 Introduction

SNP genotyping arrays have been shown to be more efficient and precise than genotyping microsatellite markers in genomewide linkage searches for Mendelian disease loci (Sellick et al., 2004). It was therefore decided to genotype the whole-genome of ACM 2 family members using the Illumina® HumanCytoSNP-12 BeadChip. This BeadChip contains 200,000 tag SNPs and 220,000 markers that target regions of known cytogenetic importance (http://www.illumina.com/products/humancytosnp_12_dna_analysis_beadchip_kits.ilmn). This experiment would refine the ARVC-6 region and may identify other disease-associated regions for this family.

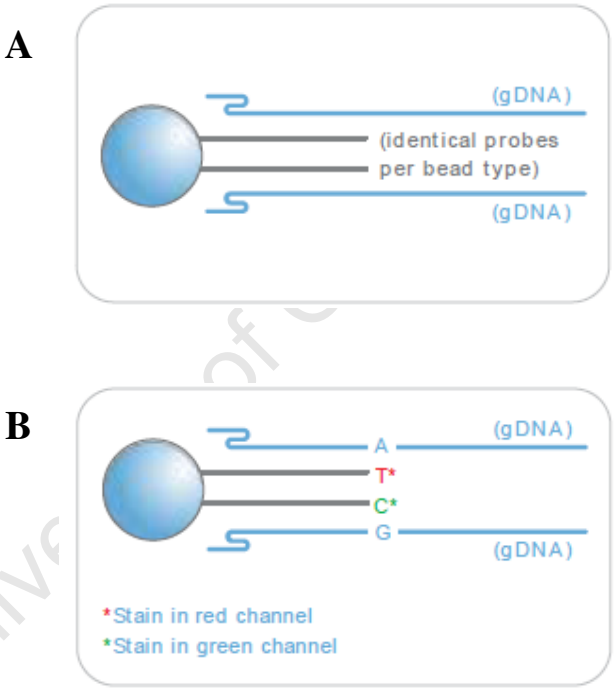
6.2 Method

6.2.1 Genotyping

A total of 14 DNA samples met the quality and quantity requirements for genotyping with the HumanCytoSNP-12 BeadChip. Since this BeadChip genotypes 12 samples at the same time, it was decided to genotype DNA samples from family members: I:2, II:3, II:4, II:6, II:8, III:1, III:5, III:6, III:8, III:9, III:11, III:12 (5 unaffected and 7 affected family members). The DNA samples were diluted to a working concentration of 50ng/µl and were sent to Xenon Pharmaceuticals Inc., Canada for genotyping on the BeadChip using the Infinium HD Assay Ultra (Infinium® HD Assay Ultra Protocol Guide).

The DNA samples were denatured, neutralised and isothermally amplified overnight. The DNA was fragmented by a controlled enzymatic process that uses end-point fragmentation to avoid overfragmenting the sample. The fragmented DNA underwent isopropanol precipitation and was subsequently collected by centrifugation at 4°C. The precipitated DNA was resuspended in hybridisation buffer. The BeadChip was prepared for hybridisation in a capillary flow-through chamber. The samples were applied to the

BeadChip and divided by an IntelliHyb® seal (or gasket). The loaded BeadChip was incubated overnight in the Illumina Hybridization Oven. During hybridisation the fragmented DNA samples annealed to locus-specific 50-mers covalently linked to the beads (see Figure 6-1A). The BeadChip was washed to remove unhybridised and non-specifically hybridised DNA. The captured DNA was used as a template to extend the oligos on the beads by a single-base (see Figure 6-1B). The Illumina iScan was used to scan the BeadChip. The scanner used a laser to excite the fluorophore of the single-base extension product on the beads. It then recorded high-resolution images of the light emitted from the fluorophores.



(Infinium® HD Assay Ultra Protocol Guide)

Figure 6-1: A. Fragmented DNA (gDNA) annealed to oligos covalently linked to a bead. **B.** Single-base extension of the oligos on a bead.

6.2.2 Analysis

Two point and multipoint linkage analyses were performed by Xenon Pharmaceuticals Inc. using Merlin (version 1.1.2) (Abecasis et al., 2002). Merlin is a computer program

that implements the Lander-Green algorithm. This program is particularly good for analysing whole-genome searches of modest sized pedigrees (Strachan and Read, 2004b) such as is the case in this study. The program parameters were set to autosomal dominant inheritance with a penetrance of 0.90, phenocopy of 0.02 and a disease allele frequency of 0.001.

Xenon Pharmaceuticals Inc. also performed haplotype analysis using 18 informative SNPs on chromosome 4 to confirm the segregation of a susceptibility allele with ARVC in family ACM 2.

6.3 Results

6.3.1 Linkage analysis

In the two-point linkage analysis lod scores above two were obtained on chromosome 1, 4, 10 and 12. Table 6-1 gives the SNP(s) with the highest two-point lod score on each of the four chromosomes. In multipoint linkage analysis lod scores above two were obtained on chromosome 4 and 10. Both regions had a peaking lod score of 2.79. The chromosome 4 -1lod interval is between pter and SNP rs13115900 which corresponds to an interval of ~6.93 Mb. The chromosome 10 -1lod interval is between SNPs rs7893899 and rs10795384 which corresponds to an interval of ~2.69 Mb.

Table 6-1: SNPs with the highest lod scores in two-point linkage analysis

Chromosome	lod	SNP
1	2.73	rs10912216
4	2.78	rs6810785 rs34782960
10	2.72	rs1155200
12	2.56	rs7977471 rs4421782 rs7978485 rs4759824

6.3.2 Haplotype analysis

In Figure 6-2 a truncated ACM 2 pedigree is shown along with the genotypes for 18 informative SNPs on chromosome 4. Affected individual III:6 has a recombination event between SNPs rs4689562 and rs10155062. The chromosome 4 linkage interval is therefore between pter and SNP rs10155062 (~7.18Mb). This is close to the -1lod interval which predicted the linkage interval between pter and SNP rs13115900.

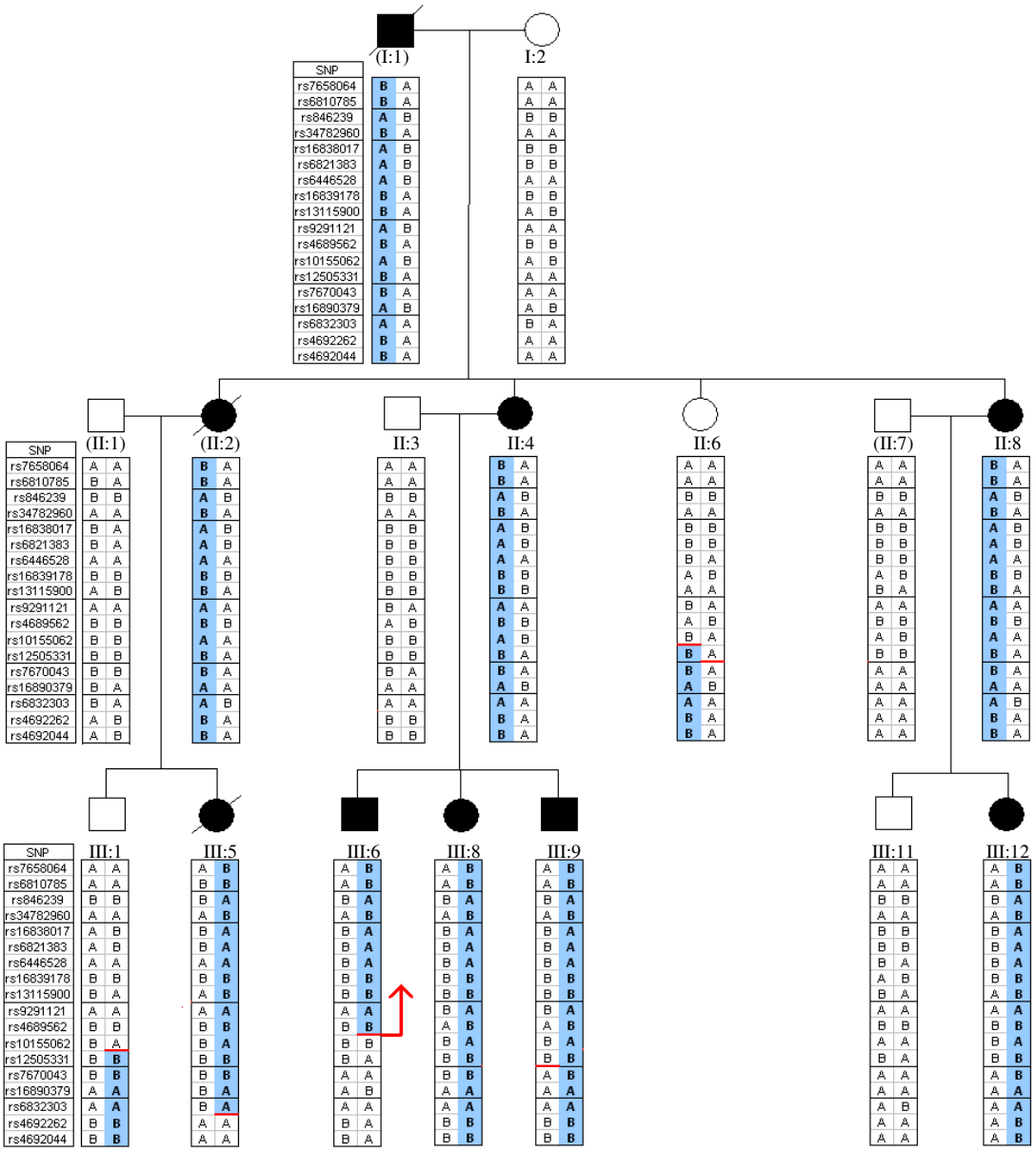


Figure 6-2: Truncated pedigree of ACM 2, showing genotypes useful for delimiting crossover points on chromosome 4. SNP names are shown at the left of each generation. Genotypes for I:1, II:1, II:2 and II:7 (pedigree numbers in brackets) are inferred. Blue bars represent the disease-associated haplotype and recombination events are marked with red lines. Individual III:6 has a recombination that delimits the region.

6.4 Discussion

A peak multipoint lod score of 2.79 was obtained on chromosome 4 and 10. The chromosome 10 -1lod interval is ~2.69Mb and correlates with the published two-point lod score of 2.93 and linkage interval of ~2.9Mb for this family (Matolweni et al., 2006). Haplotype analysis of the chromosome 4 region identified a linkage interval of ~7.18Mb. This linkage interval is not a known ARVC locus and was not previously linked to this family because linkage analysis was only performed on known ARVC loci.

This is the first time to our knowledge that whole-genome SNP linkage analysis has been used to identify a new disease locus for ARVC. The previous ARVC loci were identified by whole-genome genetic linkage analysis genotyping families with microsatellite markers or disease genes were identified by screening biologically plausible genes for mutations.

High-density genome-wide SNP arrays have assisted in identifying the disease-causing mutation in a family with ARVC and another family with DCM (Van Der Zwaag et al., 2010). Van der Zwaag and colleagues hypothesised that the longest shared haplotype, among affected family members, is most likely to contain the disease-causing mutation. They state that "short haplotypes are common and either represent ancient identical-by-descent (IBD) stretches of DNA or are identical-by-state (IBS)". The largest haplotype in the ARVC and DCM families contained known disease genes (*PKP2* and *MYH7* respectively) and mutation screening indentified the disease-causing mutation. Van der Zwaag and colleagues went on to calculate that a pedigree with at least seven meioses has a high chance of correctly detecting the mutation-containing haplotype as the largest.

In family ACM 2, the linkage interval on chromosome 4 is the largest. The linkage interval on chromosome 10 is probably shared due to hidden distant ancestry or IBS. The chromosome 4 linkage interval is therefore another plausible locus for this family. A total of 90 genes with a HGNC symbol were identified in this linkage interval using BioMart on the Ensembl database (Appendix G). These genes will need to be screened for mutations.

7 DISCUSSION

7.1 The aim of this project

The aim of this project was to identify a novel disease gene that causes ARVC in a South African family of European descent (family ACM 2).

7.2 Summary of work

Linkage analysis of family ACM 2 was previously performed at six known ARVC loci (Matolweni et al., 2006). ARVC loci 1 to 5 were excluded as disease loci in this family based on lack of evidence for linkage. However, a peak lod score of 2.93 was obtained for the ARVC-6 locus which is highly suggestive of linkage. When this project was started, another locus (ARVC-7) and five ARVC disease genes (i.e. desmoplakin, plakophilin-2, desmoglein-2, desmocollin-2 and plakoglobin) had since been reported.

Exclusion mapping was performed in family ACM 2 at the ARVC-7 locus and for five ARVC disease genes. Lod scores less than -2 were obtained for the ARVC-7 locus, desmoplakin, desmoglein-2, desmocollin-2 and plakoglobin. The lod score values obtained for plakophilin-2 do not support linkage of this locus with disease in this family. No haplotype or genotype (as is the case for desmoglein-2 and plakoglobin) segregated with ARVC in this family. The ARVC-7 locus and the five genes were therefore excluded from causing disease in this family.

Furthermore, an affected ACM 2 family member was included in a screen of South African ARVC patients for mutations in the five disease genes: plakophilin-2 (Watkins et al., 2009), desmoplakin (Fish, 2010), desmoglein-2, desmocollin-2 and plakoglobin (Mr Mzwandile Mbele, unpublished work). No disease-causing mutations were identified in this family member; this complementary approach supports the findings of the exclusion mapping of these five genes for this family.

Based on the findings of the exclusion mapping, the ARVC-6 locus was further investigated. Several existing bioinformatics tools and tissue expression data were used to prioritise positional candidate genes in the ARVC-6 locus for mutation screening.

Comprehensive mutation screening was conducted for 15 genes using different methods. A total of 36 sequence variants (SVs) were identified. Twenty-nine SVs have been reported as SNPs on the NCBI SNP database and seven SVs are novel. Only one SV, c.1641A>G in *DCLRE1C*, was found in the two affected family members and was not present in the two unaffected family members screened. This substitution results in a new codon but it specifies the same amino acid (i.e. synonymous substitution). The significance of this SV is unknown.

While no overt disease-causing mutations were identified in the coding regions and intron-exon boundaries, mutations in the 5' and 3' untranslated regions (UTRs) cannot be ruled out. Beffagna and colleagues reported disease-causing mutations in the 5' and 3' UTRs of the *TGFβ3* gene (Beffagna et al., 2005). In a canine model of ARVC, dogs homozygous for an 8bp deletion in the 3' UTR of the striatin gene had a more severe form of disease (Meurs et al., 2010).

The mutation screening methods employed in this study (i.e. sequencing, High Resolution Melt (HRM) and Denaturing High Performance Liquid Chromatography (DHPLC)) would not have identified copy-number variations (CNVs). Copy-number analysis, of one affected family member, was performed using an Affymetrix® Genome-Wide Human SNP Array 6.0, the Affymetrix Genotyping Console® 4.0 and a reference model file provided by Affymetrix. No CNV was found in the ARVC-6 region.

Whole-genome SNP linkage analysis was performed in family ACM 2 to refine the ARVC-6 region and possibly identify other disease-associated regions. A peak multipoint lod score of 2.79 was obtained on chromosome 4 and 10. The chromosome 10 -1lod interval is ~2.69Mb and correlates with the published two-point lod score of 2.93 and linkage interval of ~2.9Mb for this family (Matolweni et al., 2006). Haplotype analysis of the chromosome 4 region identified a linkage interval of ~7.18Mb. This linkage interval

is not a known ARVC locus and was not previously linked to this family because linkage analysis was only performed on known ARVC loci.

This is the first time to our knowledge that whole-genome SNP linkage analysis has been used to identify a new disease locus for ARVC. The previous ARVC loci were identified by whole-genome genetic linkage analysis genotyping families with microsatellite markers or disease genes were identified by screening biologically plausible genes for mutations.

High-density genome-wide SNP arrays have assisted in identifying the disease-causing mutation in a family with ARVC and another family with DCM (Van Der Zwaag et al., 2010). Van der Zwaag and colleagues hypothesised that the longest shared haplotype, among affected family members, is most likely to contain the disease-causing mutation. They state that "short haplotypes are common and either represent ancient identical-by-descent (IBD) stretches of DNA or are identical-by-state (IBS)". The largest haplotype in the ARVC and DCM families contained known disease genes (*PKP2* and *MYH7* respectively) and mutation screening indentified the disease-causing mutation. Van der Zwaag and colleagues went on to calculate that a pedigree with at least seven meioses has a high chance of correctly detecting the mutation-containing haplotype as the largest.

In family ACM 2, the linkage interval on chromosome 4 is the largest. The linkage interval on chromosome 10 is probably shared due to hidden distant ancestry or IBS. The chromosome 4 linkage interval is therefore another plausible locus for this family.

7.3 Future work

It is important to screen the synonynous substitution found in *DCLRE1C* in the rest of the ACM 2 family. If the SV is found to track with disease in the family then its effect on pre-mRNA splicing will need to be experimentally determined.

A total of 90 genes with a HGNC symbol were identified in the chromosome 4 linkage interval using BioMart on the Ensembl database. These genes could be prioritised and screened for mutations or the whole linkage interval could be sequenced using next-generation sequencing (NGS) to identify all variation in the region.

REFERENCES

Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet *30*, 97-101.

Adie, E.A., Adams, R.R., Evans, K.L., Porteous, D.J., and Pickard, B.S. (2005). Speeding disease gene discovery by sequence based candidate prioritization. BMC Bioinformatics *6*, 55.

Adie, E.A., Adams, R.R., Evans, K.L., Porteous, D.J., and Pickard, B.S. (2006). SUSPECTS: enabling fast and effective prioritization of positional candidates. Bioinformatics *22*, 773-774.

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.C., De Moor, B., Marynen, P., Hassan, B.*, et al.* (2006). Gene prioritization through genomic data fusion. Nat Biotechnol *24*, 537-544.

Ahmad, F., Li, D., Karibe, A., Gonzalez, O., Tapscott, T., Hill, R., Weilbaecher, D., Blackie, P., Furey, M., Gardner, M.*, et al.* (1998). Localization of a gene responsible for arrhythmogenic right ventricular dysplasia to chromosome 3p23. Circulation *98*, 2791-2795.

Alcalai, R., Metzger, S., Rosenheck, S., Meiner, V., and Chajek-Shaul, T. (2003). A recessive mutation in desmoplakin causes arrhythmogenic right ventricular dysplasia, skin disorder, and woolly hair. J Am Coll Cardiol *42*, 319-327.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Asimaki, A., Syrris, P., Wichter, T., Matthias, P., Saffitz, J.E., and McKenna, W.J. (2007). A novel dominant mutation in plakoglobin causes arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet *81*, 964-973.

Awad, M.M., Dalal, D., Tichnell, C., James, C., Tucker, A., Abraham, T., Spevak, P.J., Calkins, H., and Judge, D.P. (2006). Recessive arrhythmogenic right ventricular dysplasia due to novel cryptic splice mutation in PKP2. Hum Mutat *27*, 1157.

Bauce, B., Nava, A., Beffagna, G., Basso, C., Lorenzon, A., Smaniotto, G., De Bortoli, M., Rigato, I., Mazzotti, E., Steriotis, A.*, et al.* (2010). Multiple mutations in desmosomal proteins encoding genes in arrhythmogenic right ventricular cardiomyopathy/dysplasia. Heart Rhythm *7*, 22-29.

Beffagna, G., Occhi, G., Nava, A., Vitiello, L., Ditadi, A., Basso, C., Bauce, B., Carraro, G., Thiene, G., Towbin, J.A.*, et al.* (2005). Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. Cardiovasc Res *65*, 366-373.

Bradshaw, D., Groenewald, P., Laubscher, R., and Nannan, N. (2003). Initial estimates from the South African National Burden of Disease Study, 2000. In MRC Policy Brief No1 (Cape Town, MRC).

Campbell, R.M. (2005). The treatment of cardiac causes of sudden death, syncope, and seizure. Semin Pediatr Neurol *12*, 59-66.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Chelu, M.G., Danila, C.I., Gilman, C.P., and Hamilton, S.L. (2004). Regulation of ryanodine receptors by FK506 binding proteins. Trends Cardiovasc Med *14*, 227-234.

Coonar, A.S., Protonotarios, N., Tsatsopoulou, A., Needham, E.W., Houlston, R.S., Cliff, S., Otter, M.I., Murday, V.A., Mattu, R.K., and McKenna, W.J. (1998). Gene for arrhythmogenic right ventricular cardiomyopathy with diffuse nonepidermolytic

palmoplantar keratoderma and woolly hair (Naxos disease) maps to 17q21. Circulation *97*, 2049-2058.
Corrado, D., Basso, C., Rizzoli, G., Schiavon, M., and Thiene, G. (2003). Does sports activity enhance the risk of sudden death in adolescents and young adults? J Am Coll Cardiol *42*, 1959-1963.
Corrado, D., Thiene, G., Nava, A., Rossi, L., and Pennelli, N. (1990). Sudden death in young competitive athletes: clinicopathologic correlations in 22 cases. Am J Med *89*, 588-596.
Dalla Volta, S., Battaglia, G., and Zerbini, E. (1961). "Auricularization' of right ventricular pressure curve. Am Heart J *61*, 25-33.
Dokuparti, M.V., Pamuru, P.R., Thakkar, B., Tanjore, R.R., and Nallari, P. (2005). Etiopathogenesis of arrhythmogenic right ventricular cardiomyopathy. J Hum Genet *50*, 375-381.
Elliott, P., Andersson, B., Arbustini, E., Bilinska, Z., Cecchi, F., Charron, P., Dubourg, O., Kuhl, U., Maisch, B., McKenna, W.J.*, et al.* (2008). Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. Eur Heart J *29*, 270-276.
Fish, M. (2010). Mutation Analysis of the Desmoplakin Gene in Cardiomyopathy. In Medicine (University of Cape Town).
Frances, R., Rodriguez Benitez, A.M., and Cohen, D.R. (1997). Arrhythmogenic right ventricular dysplasia and anterior polar cataract. Am J Med Genet *73*, 125-126.
Frias, P.A. (2005). Hypertrophic cardiomyopathy and arrhythmogenic right ventricular dysplasia in young patients. Semin Pediatr Neurol *12*, 25-31.
Gerull, B., Heuser, A., Wichter, T., Paul, M., Basson, C.T., McDermott, D.A., Lerman, B.B., Markowitz, S.M., Ellinor, P.T., MacRae, C.A.*, et al.* (2004). Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. Nat Genet *36*, 1162-1164.
Herren, T., Gerber, P.A., and Duru, F. (2009). Arrhythmogenic right ventricular cardiomyopathy/dysplasia: a not so rare "disease of the desmosome" with multiple clinical presentations. Clin Res Cardiol *98*, 141-158.
Heuser, A., Plovie, E.R., Ellinor, P.T., Grossmann, K.S., Shin, J.T., Wichter, T., Basson, C.T., Lerman, B.B., Sasse-Klaassen, S., Thierfelder, L.*, et al.* (2006). Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet *79*, 1081-1088.
Hodgkinson, K.A., Parfrey, P.S., Bassett, A.S., Kupprion, C., Drenckhahn, J., Norman, M.W., Thierfelder, L., Stuckless, S.N., Dicks, E.L., McKenna, W.J.*, et al.* (2005). The impact of implantable cardioverter-defibrillator therapy on survival in autosomal-dominant arrhythmogenic right ventricular cardiomyopathy (ARVD5). J Am Coll Cardiol *45*, 400-408.
Huang, Q.Y., Li, G.H., Cheung, W.M., Song, Y.Q., and Kung, A.W. (2008). Prediction of osteoporosis candidate genes by computational disease-gene identification strategy. J Hum Genet *53*, 644-655.
Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L.*, et al.* (2009). InterPro: the integrative protein signature database. Nucleic Acids Res *37*, D211-215.
John, A.S., Mohiaddin, R.H., and Sheppard, M.N. (2004). Arrhythmogenic right ventricular cardiomyopathy. Heart *90*, 1102.
Laitinen, P.J., Brown, K.M., Piippo, K., Swan, H., Devaney, J.M., Brahmbhatt, B., Donarum, E.A., Marino, M., Tiso, N., Viitasalo, M.*, et al.* (2001). Mutations of the cardiac

ryanodine receptor (RyR2) gene in familial polymorphic ventricular tachycardia. Circulation *103*, 485-490.
Lal, A., Lash, A.E., Altschul, S.F., Velculescu, V., Zhang, L., McLendon, R.E., Marra, M.A., Prange, C., Morin, P.J., Polyak, K., *et al.* (1999). A public database for gene expression in human cancers. Cancer Res *59*, 5403-5407.
Latib, M.A., Michael, K.A., and Mayosi, B.M. (2004). Initial Report of the Arrhythmogenic Right Ventricular Cardiomyopathy Registry of South Africa: 2004 [Abstract]. Cardiovascular Journal of South Africa *15*, 237 & 238.
Li, D., Ahmad, F., Gardner, M.J., Weilbaecher, D., Hill, R., Karibe, A., Gonzalez, O., Tapscott, T., Sharratt, G.P., Bachinski, L.L., *et al.* (2000a). The locus of a novel gene responsible for arrhythmogenic right-ventricular dysplasia characterized by early onset and high penetrance maps to chromosome 10p12-p14. Am J Hum Genet *66*, 148-156.
Li, D., Bachinski, L.L., and Roberts, R. (2001). Genomic organization and isoform-specific tissue expression of human NAPOR (CUGBP2) as a candidate gene for familial arrhythmogenic right ventricular dysplasia. Genomics *74*, 396-401.
Li, D., Gonzalez, O., Bachinski, L.L., and Roberts, R. (2000b). Human protein tyrosine phosphatase-like gene: expression profile, genomic structure, and mutation analysis in families with ARVD. Gene *256*, 237-243.
Lombard, Z., Tiffin, N., Hofmann, O., Bajic, V.B., Hide, W., and Ramsay, M. (2007). Computational selection and prioritization of candidate genes for fetal alcohol syndrome. BMC Genomics *8*, 389.
Lopez-Bigas, N., and Ouzounis, C.A. (2004). Genome-wide identification of genes likely to be involved in human genetic disease. Nucleic Acids Res *32*, 3108-3114.
Marcus, F.I., McKenna, W.J., Sherrill, D., Basso, C., Bauce, B., Bluemke, D.A., Calkins, H., Corrado, D., Cox, M.G., Daubert, J.P., *et al.* (2010). Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the Task Force Criteria. Eur Heart J *31*, 806-814.
Marks, A.R. (2000). Cardiac intracellular calcium release channels: role in heart failure. Circ Res *87*, 8-11.
Maron, B.J., Shirani, J., Poliac, L.C., Mathenge, R., Roberts, W.C., and Mueller, F.O. (1996). Sudden death in young competitive athletes. Clinical, demographic, and pathological profiles. Jama *276*, 199-204.
Matolweni, L.O., Bardien, S., Rebello, G., Oppon, E., Munclinger, M., Ramesar, R., Watkins, H., and Mayosi, B.M. (2006). Arrhythmogenic right ventricular cardiomyopathy type 6 (ARVC6): support for the locus assignment, narrowing of the critical region and mutation screening of three candidate genes. BMC Med Genet *7*, 29.
Mayosi, B.M. (2007). Contemporary trends in the epidemiology and management of cardiomyopathy and pericarditis in sub-Saharan Africa. Heart *93*, 1176-1183.
McGrath, J.A. (2005). Inherited disorders of desmosomes. Australas J Dermatol *46*, 221-229.
McKenna, W.J., Thiene, G., Nava, A., Fontaliran, F., Blomstrom-Lundqvist, C., Fontaine, G., and Camerini, F. (1994). Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology. Br Heart J *71*, 215-218.
McKoy, G., Protonotarios, N., Crosby, A., Tsatsopoulou, A., Anastasakis, A., Coonar, A., Norman, M., Baboonian, C., Jeffery, S., and McKenna, W.J. (2000). Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). Lancet *355*, 2119-2124.

Melberg, A., Oldfors, A., Blomstrom-Lundqvist, C., Stalberg, E., Carlsson, B., Larrson, E., Lidell, C., Eeg-Olofsson, K.E., Wikstrom, G., Henriksson, G.*, et al.* (1999). Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy linked to chromosome 10q. Ann Neurol *46*, 684-692.
Mendez, G.F., and Cowie, M.R. (2001). The epidemiological features of heart failure in developing countries: a review of the literature. Int J Cardiol *80*, 213-219.
Merner, N.D., Hodgkinson, K.A., Haywood, A.F., Connors, S., French, V.M., Drenckhahn, J.D., Kupprion, C., Ramadanova, K., Thierfelder, L., McKenna, W.*, et al.* (2008). Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. Am J Hum Genet *82*, 809-821.
Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat Rev Genet *11*, 31-46.
Meurs, K.M., Mauceli, E., Lahmers, S., Acland, G.M., White, S.N., and Lindblad-Toh, K. (2010). Genome-wide association identifies a deletion in the 3' untranslated region of striatin in a canine model of arrhythmogenic right ventricular cardiomyopathy. Hum Genet *128*, 315-324.
Michael, K.A., Latib, M.A., Matolweni, L.O., and Mayosi, B.M. (2004). Arrhythmogenic right ventricular cardiomyopathy: pathogenesis, diagnosis and management. SA Heart *1*, 13-22.
Munclinger, M.J., Patel, J.J., and Mitha, A.S. (2000). Follow-up of patients with arrhythmogenic right ventricular cardiomyopathy dysplasia. S Afr Med J *90*, 61-68.
Nattel, S., and Schott, J.J. (2005). Arrhythmogenic right ventricular dysplasia type 1 and mutations in transforming growth factor beta3 gene regulatory regions: a breakthrough? Cardiovasc Res *65*, 302-304.
Oefner, P.J., and Underhill, P.A. (1995). Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC). American Journal of Human Genetics *57*, A266.
Oefner, P.J., and Underhill, P.A. (1999). DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). In Current Protocols in Human Genetics (New York, Wiley-Interscience), pp. 1 - 12.
Paul, M., Schulze-Bahr, E., Breithardt, G., and Wichter, T. (2003). Genetics of arrhythmogenic right ventricular cardiomyopathy--status quo and future perspectives. Z Kardiol *92*, 128-136.
Perez-Iratxeta, C., Bork, P., and Andrade-Navarro, M.A. (2007). Update of the G2D tool for prioritization of gene candidates to inherited diseases. Nucleic Acids Res *35*, W212-216.
Peters, S., Trummel, M., and Meyners, W. (2004). Prevalence of right ventricular dysplasia-cardiomyopathy in a non-referral hospital. Int J Cardiol *97*, 499-501.
Pilichou, K., Nava, A., Basso, C., Beffagna, G., Bauce, B., Lorenzon, A., Frigo, G., Vettori, A., Valente, M., Towbin, J.*, et al.* (2006). Mutations in Desmoglein-2 Gene Are Associated With Arrhythmogenic Right Ventricular Cardiomyopathy. Circulation.
Priori, S.G., Napolitano, C., Tiso, N., Memmi, M., Vignati, G., Bloise, R., Sorrentino, V., and Danieli, G.A. (2001). Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie catecholaminergic polymorphic ventricular tachycardia. Circulation *103*, 196-200.
Protonotarios, N., Tsatsopoulou, A., Anastasakis, A., Sevdalis, E., McKoy, G., Stratos, K., Gatzoulis, K., Tentolouris, K., Spiliopoulou, C., Panagiotakos, D.*, et al.* (2001). Genotype-phenotype assessment in autosomal recessive arrhythmogenic right ventricular

cardiomyopathy (Naxos disease) caused by a deletion in plakoglobin. J Am Coll Cardiol *38*, 1477-1484.
Protonotarios, N., Tsatsopoulou, A., Patsourakos, P., Alexopoulos, D., Gezerlis, P., Simitsis, S., and Scampardonis, G. (1986). Cardiac abnormalities in familial palmoplantar keratosis. Br Heart J *56*, 321-326.
Protonotarios, N.I., Tsatsopoulou, A.A., and Gatzoulis, K.A. (2002). Arrhythmogenic right ventricular cardiomyopathy caused by a deletion in plakoglobin (Naxos disease). Card Electrophysiol Rev *6*, 72-80.
Rampazzo, A., Nava, A., Danieli, G.A., Buja, G., Daliento, L., Fasoli, G., Scognamiglio, R., Corrado, D., and Thiene, G. (1994). The gene for arrhythmogenic right ventricular cardiomyopathy maps to chromosome 14q23-q24. Hum Mol Genet *3*, 959-962.
Rampazzo, A., Nava, A., Erne, P., Eberhard, M., Vian, E., Slomp, P., Tiso, N., Thiene, G., and Danieli, G.A. (1995). A new locus for arrhythmogenic right ventricular cardiomyopathy (ARVD2) maps to chromosome 1q42-q43. Hum Mol Genet *4*, 2151-2154.
Rampazzo, A., Nava, A., Malacrida, S., Beffagna, G., Bauce, B., Rossi, V., Zimbello, R., Simionati, B., Basso, C., Thiene, G.*, et al.* (2002). Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. Am J Hum Genet *71*, 1200-1206.
Rampazzo, A., Nava, A., Miorin, M., Fonderico, P., Pope, B., Tiso, N., Livolsi, B., Zimbello, R., Thiene, G., and Danieli, G.A. (1997). ARVD4, a new locus for arrhythmogenic right ventricular cardiomyopathy, maps to chromosome 2 long arm. Genomics *45*, 259-263.
Reddy, S., Jia, S., Geoffrey, R., Lorier, R., Suchi, M., Broeckel, U., Hessner, M.J., and Verbsky, J. (2009). An autoinflammatory disease due to homozygous deletion of the IL1RN locus. N Engl J Med *360*, 2438-2444.
Richardson, P., McKenna, W., Bristow, M., Maisch, B., Mautner, B., O'Connell, J., Olsen, E., Thiene, G., Goodwin, J., Gyarfas, I.*, et al.* (1996). Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of cardiomyopathies. Circulation *93*, 841-842.
Sellick, G.S., Longman, C., Tolmie, J., Newbury-Ecob, R., Geenhalgh, L., Hughes, S., Whiteford, M., Garrett, C., and Houlston, R.S. (2004). Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. Nucleic Acids Res *32*, e164.
Sen-Chowdhry, S., Morgan, R.D., Chambers, J.C., and McKenna, W.J. (2010). Arrhythmogenic cardiomyopathy: etiology, diagnosis, and treatment. Annu Rev Med *61*, 233-253.
Sen-Chowdhry, S., Syrris, P., and McKenna, W.J. (2007a). Role of genetic analysis in the management of patients with arrhythmogenic right ventricular dysplasia/cardiomyopathy. J Am Coll Cardiol *50*, 1813-1821.
Sen-Chowdhry, S., Syrris, P., Ward, D., Asimaki, A., Sevdalis, E., and McKenna, W.J. (2007b). Clinical and genetic characterization of families with arrhythmogenic right ventricular dysplasia/cardiomyopathy provides novel insights into patterns of disease expression. Circulation *115*, 1710-1720.
Severini, G.M., Krajinovic, M., Pinamonti, B., Sinagra, G., Fioretti, P., Brunazzi, M.C., Falaschi, A., Camerini, F., Giacca, M., and Mestroni, L. (1996). A new locus for arrhythmogenic right ventricular dysplasia on the long arm of chromosome 14. Genomics *31*, 193-200.

Shmueli, O., Horn-Saban, S., Chalifa-Caspi, V., Shmoish, M., Ophir, R., Benjamin-Rodrig, H., Safran, M., Domany, E., and Lancet, D. (2003). GeneNote: whole genome expression profiles in normal human tissues. C R Biol *326*, 1067-1072.
Sliwa, K., Damasceno, A., and Mayosi, B.M. (2005). Epidemiology and etiology of cardiomyopathy in Africa. Circulation *112*, 3577-3583.
Sliwa, K., Wilkinson, D., Hansen, C., Ntyintyane, L., Tibazarwa, K., Becker, A., and Stewart, S. (2008). Spectrum of heart disease and risk factors in a black urban population in South Africa (the Heart of Soweto Study): a cohort study. Lancet *371*, 915-922.
Stankiewicz, P., and Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. Annu Rev Med *61*, 437-455.
Strachan, T., and Read, A.P. (2004a). Chapter 13: Genetic mapping of Mendelian characters. In Human Molecular Genetics 3 (London and New York, Garland Science), pp. 398, 406.
Strachan, T., and Read, A.P. (2004b). Chapter 13: Genetic mapping of Mendelian characters. In Human Molecular Genetics 3 (London and New York, Garland Science), p. 412.
Syrris, P., Ward, D., Evans, A., Asimaki, A., Gandjbakhch, E., Sen-Chowdhry, S., and McKenna, W.J. (2006). Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. Am J Hum Genet *79*, 978-984.
Thiel, C.T., Horn, D., Zabel, B., Ekici, A.B., Salinas, K., Gebhart, E., Ruschendorf, F., Sticht, H., Spranger, J., Muller, D.*, et al.* (2005). Severely incapacitating mutations in patients with extreme short stature identify RNA-processing endoribonuclease RMRP as an essential cell growth regulator. Am J Hum Genet *77*, 795-806.
Thiene, G., Corrado, D., Nava, A., Rossi, L., Poletti, A., Boffa, G.M., Daliento, L., and Pennelli, N. (1991). Right ventricular cardiomyopathy: is there evidence of an inflammatory aetiology? Eur Heart J *12 Suppl D*, 22-25.
Thiene, G., Nava, A., Corrado, D., Rossi, L., and Pennelli, N. (1988). Right ventricular cardiomyopathy and sudden death in young people. N Engl J Med *318*, 129-133.
Tiffin, N., Adie, E., Turner, F., Brunner, H.G., van Driel, M.A., Oti, M., Lopez-Bigas, N., Ouzounis, C., Perez-Iratxeta, C., Andrade-Navarro, M.A.*, et al.* (2006). Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. Nucleic Acids Res *34*, 3067-3081.
Tiffin, N., Andrade-Navarro, M.A., and Perez-Iratxeta, C. (2009). Linking genes to diseases: it's all in the data. Genome Med *1*, 77.
Tiffin, N., Kelso, J.F., Powell, A.R., Pan, H., Bajic, V.B., and Hide, W.A. (2005). Integration of text- and data-mining using ontologies successfully selects disease gene candidates. Nucleic Acids Res *33*, 1544-1552.
Tiffin, N., Okpechi, I., Perez-Iratxeta, C., Andrade-Navarro, M.A., and Ramesar, R. (2008). Prioritization of candidate disease genes for metabolic syndrome by computational analysis of its defining phenotypes. Physiol Genomics *35*, 55-64.
Tiso, N., Stephan, D.A., Nava, A., Bagattin, A., Devaney, J.M., Stanchi, F., Larderet, G., Brahmbhatt, B., Brown, K., Bauce, B.*, et al.* (2001). Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). Hum Mol Genet *10*, 189-194.
Tremblay, K., Lemire, M., Potvin, C., Tremblay, A., Hunninghake, G.M., Raby, B.A., Hudson, T.J., Perez-Iratxeta, C., Andrade-Navarro, M.A., and Laprise, C. (2008). Genes to diseases (G2D) computational method to identify asthma candidate genes. PLoS One *3*, e2907.

Van Der Zwaag, P., Van Tintelen, J., Gerbens, F., Jongbloed, J., Boven, L., Van Der Smagt, J., Van Der Roest, W., Van Langen, I., Bikker, H., Hauer, R.*, et al.* (2010). Haplotype sharing test maps genes for familial cardiomyopathies. Clin Genet *[Epub ahead of print]*.
van Driel, M.A., Cuelenaere, K., Kemmeren, P.P., Leunissen, J.A., and Brunner, H.G. (2003). A new web-based data mining tool for the identification of candidate genes for human genetic disorders. Eur J Hum Genet *11*, 57-63.
van Driel, M.A., Cuelenaere, K., Kemmeren, P.P., Leunissen, J.A., Brunner, H.G., and Vriend, G. (2005). GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases. Nucleic Acids Res *33*, W758-761.
van Tintelen, J.P., Van Gelder, I.C., Asimaki, A., Suurmeijer, A.J., Wiesfeld, A.C., Jongbloed, J.D., van den Wijngaard, A., Kuks, J.B., van Spaendonck-Zwarts, K.Y., Notermans, N.*, et al.* (2009). Severe cardiac phenotype with right ventricular predominance in a large cohort of patients with a single missense mutation in the DES gene. Heart Rhythm *6*, 1574-1583.
Watkins, D.A., Hendricks, N., Shaboodien, G., Mbele, M., Parker, M., Vezi, B.Z., Latib, A., Chin, A., Little, F., Badri, M.*, et al.* (2009). Clinical features, survival experience, and profile of plakophylin-2 gene mutations in participants of the arrhythmogenic right ventricular cardiomyopathy registry of South Africa. Heart Rhythm *6*, S10-17.
Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A.*, et al.* (2003). Database resources of the National Center for Biotechnology. Nucleic Acids Res *31*, 28-33.
Wichter, T., and Breithardt, G. (2005). Implantable cardioverter-defibrillator therapy in arrhythmogenic right ventricular cardiomyopathy: a role for genotyping in decision-making? J Am Coll Cardiol *45*, 409-411.
Williams, R.G., and Chen, A.Y. (2003). Identifying athletes at risk for sudden death. J Am Coll Cardiol *42*, 1964-1966.
Wolfe, B.A., and Corwin, E.J. (2005). ARVD: a silent killer. Nurse Pract *30*, 16-18, 23-17; quiz 27-19.
Xiao, W., and Oefner, P.J. (2001). Denaturing high-performance liquid chromatography: A review. Hum Mutat *17*, 439-474.
Xu, T., Yang, Z., Vatta, M., Rampazzo, A., Beffagna, G., Pilichou, K., Scherer, S.E., Saffitz, J., Kravitz, J., Zareba, W.*, et al.* (2010). Compound and digenic heterozygosity contributes to arrhythmogenic right ventricular cardiomyopathy. J Am Coll Cardiol *55*, 587-597.
Yin, T., and Green, K.J. (2004). Regulation of desmosome assembly and adhesion. Semin Cell Dev Biol *15*, 665-677.

WEB RESOURCES

Affymetrix® FAQs

http://www.affymetrix.com/support/help/faqs/gw_human_snp6/faq_10.jsp

Affymetrix® Genome-Wide Human SNP Array 6.0 Data Sheet

http://media.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf

BioEdit

http://www.mbio.ncsu.edu/bioedit/bioedit.html

BioMart

http://www.ensembl.org/biomart/index.html

BLAST

http://blast.ncbi.nlm.nih.gov/

Disease Gene Prediction (DGP)

http://maine.ebi.ac.uk:8000/services/dgp

Dragon Disease Explorer (DDE)

http://research.i2r.a-star.edu.sg/DRAGON/DE/

Endeavour

http://www.esat.kuleuven.be/endeavour

Ensembl database

http://www.ensembl.org/index.html

e-PCR

http://www.ncbi.nlm.nih.gov/sutils/e-pcr

The Genome Database

http://www.gdb.org

G2D (genes to diseases)

http://www.ogic.ca/projects/g2d_2/

GeneCards® database

http://www.genecards.org

GeneSeeker

http://www.cmbi.ru.nl/GeneSeeker/

HRM Assay Design and Analysis booklet

http://www.corbettlifescience.com/shared/Rotor-Gene%206000/hrm_corprotocol.pdf

Illumina® HumanCytoSNP-12 BeadChip

http://www.illumina.com/products/humancytosnp_12_dna_analysis_beadchip_kits.ilmn

Infinium® HD Assay Ultra Protocol Guide

Infinium_HD_Ultra_User_Guide_11328087_RevB.pdf

NCBI

http://www.ncbi.nlm.nih.gov/

OligoAnalyzer 3.0

http://scitools.idtdna.com/scitools/Applications/OligoAnalyzer/Default.aspx

OMIM (Online Mendelian Inheritance in Man)

http://www.ncbi.nlm.nih.gov/omim/

Primer 3

http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_results.cgi

PROSPECTR

http://www.genetics.med.ed.ac.uk/prospectr/

PubMed

http://www.ncbi.nlm.nih.gov/pubmed/

Structure & Function of GeneChip® Microarrays

http://media.affymetrix.com/about_affymetrix/outreach/lesson_plan/downloads/student_manual_activities/activity2/activity2_structure_function.pdf

SUSPECTS

http://www.genetics.med.ed.ac.uk/suspects/

UniSTS database on NCBI

http://www.ncbi.nlm.nih.gov/unists/

WHO (World Health Organisation) Fact sheet no. 317

http://www.who.int/mediacentre/factsheets/fs317/en/index.html

APPENDICES

Appendix A: Consent Form

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory
Division of Human Genetics
1st Floor, Anatomy Building
UCT Medical School, Observatory 7925

Tel: (021) 406 6425 Fax: (021) 448-0906

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel. Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB
Keep blood in fridge at 4°C until able to send to laboratory

Please DO NOT send specimens on ice or frozen.

Please fill in all the information requested:

Surname: _____ First Name(s): _____

New Family: Yes ☐ No ☐ (If no, please fill in family name) Family name: _____

Medical Aid: _____ Medical Aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year: _____ Month: _____ Day: _____

Number of children: _____

Ethnic Origin : (please indicate ancestry of both your mother and father) _____

Contact Address: _____ Town: _____ Fax: Tel: _____

Referring Doctor/Sister: _____ Town: _____ Fax: Tel: _____

Hospital or Address: _____ Town: _____ Fax: Tel: _____

Reason for Referral (Clinical diagnosis):

Affected ☐ At Risk ☐ Carrier ☐ Spouse ☐ Query ☐ Unaffected ☐

Arrhythmogenic Right Ventricular Cardiomyopathy ☐

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

Physical disability ☐ Mental retardation ☐ Deafness ☐ Impaired vision ☐ Night blindness ☐

Other: _____

Have samples from this patient been sent to a DNA lab before? (DELETE WHERE NOT APPLICABLE) YES / NO / Don't Know

If Yes, where: _____

For Laboratory use only:

DNA number: _____ *Vol.Blood:* _____ *(ml)* *Other:* _____

Date Received: Year: _____ *Month:* _____ *Day:* _____ *Computer Index No:* _____

CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____ , request that an attempt be made using genetic material to assess the probability that: I might have inherited a disease-causing mutation in the gene for arrhythmogenic right ventricular cardiomyopathy.
2. I understand that the genetic material for analysis is to be obtained from: blood cells/other (specify) (DELETE WHERE NOT APPLICABLE) :
3. I request that **no** portion of the sample be stored for later use. ☐ (MARK IF APPLICABLE)
   Or
   I request that a portion of the sample be stored indefinitely for (DELETE WHERE NOT APPLICABLE):
   (a) possible re-analysis
   (b) analysis for the benefit of members of my immediate family
   (c) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee, provided that any information from such research will remain confidential.
4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available.
   In addition, I authorise that they may be made known to: (DELETE WHERE NOT APPLICABLE) :
   other doctors involved in my care
   the following family members: _____
   other: _____
5. I authorise / do not authorise my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UCT.
6. I have been informed that:
   (a) there are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   (b) the analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   (c) the genetics laboratory is under an obligation to respect medical confidentiality .
   (d) genetic analysis may not be informative for some families or family members.
   (e) even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   (f) where biological material is used for research purposes, there may be no direct benefit to me.
7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.
8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

_____ DATE: _____

Patient signature _____ **Witnessed consent** _____

NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM

Appendix B: Primers

Table B-1: Primer sequences for microsatellite markers (set one)

Locus & disease gene (if known)	Microsatellite marker	Max Het.	Type of repeat	Name of primer	Primer sequence (5'-3')	PCR product size (in bp)	Fluorescent tag	Ta (in ^{o}C)
ARVC-7	D10S1752	0.6700	CA	D10S1752_F	tactgtcctctatttcatttctaca	275-285	FAM	*a*
				D10S1752_R	aatttattacaagcaaaaccaccct		-	
	D10S2327	no info	GGAT	D10S2327_F	cccagagcaagtactcacct	200-228	FAM	52
				D10S2327_R	atagttttgtgcttattgacatga		-	
	D10S201	0.8331	CA	D10S201_F	ggcatggtttgaggatgagaggg	210	HEX	49
				D10S201_R	aaggctgctggagaacagagaaag		-	
ARVC-8 (*DSP*)	D6S1547	0.6606	CA	D6S1547_F	ccttgagcaccttaaattttt	290-304	HEX	*a*
				D6S1547_R	taactgacaaagcagaatagca		-	
ARVC-9 (*PKP2*)	D12S1692	0.8349	CA	D12S1692_F	ctttgattccataccctcct	245-261	HEX	*a*
				D12S1692_R	gcagcaatttcagacttctc		-	
ARVC-10 (*DSG2*)	D18S36	0.7500	CA	D18S36_F	ttcagttttccacatgcataaaa	146	HEX	49

				D18S36_R	tcttccatttatcccaaaatg		-	
ARVC-11 (*DSC2*)	AFM344ZD5	no info	CA	AFM344ZD5_F	gttttagtgcctgtgcc	132	FAM	52
				AFM344ZD5_R	atccaaattccagttctca		-	
Naxos disease (*JUP*)	D17S846	0.8290	GGAA	D17S846_F	tgcatacctgtactacttcag	215	FAM	*a*
				D17S846_R	tcctttgttgcagatttcttc		-	

Note: *DSP* = Desmoplakin; *PKP2* = Plakophilin-2; *DSG2* = Desmoglein-2; *DSC2* = Desmocollin-2; *JUP* = Plakoglobin; Max Het. = Maximum Heterozygosity (from The Genome Database http://www.gdb.org); F = forward primer; R = reverse primer; PCR product sizes from http://www.ncbi.nlm.nih.gov/unists/; Ta = Annealing temperature; *a* = Optimised by Michelle Parker

Table B-2: Primer sequences for microsatellite markers (set two)

Locus & disease gene	Microsatellite marker	Type of repeat	Name of primer	Primer sequence (5'-3')	PCR product size (in bp)	Fluorescent tag	Ta (in ^{o}C)
ARVC-8 (DSP)	DSP-M1	AC	DSPMarker_1F	tcgggagaattcccacatta	240	FAM	56
			DSPMarker_1R	tcctcacggatgtgctacaa		-	
	DSP-M2	GT	DSPMarker_2F	tgaagggtcaggtgtcatca	229	HEX	56
			DSPMarker_2R	atatgccagggtggcttctt		-	
ARVC-9 (PKP2)	PKP2-M1	CA	PKP2Marker_1F	ctctcaaatagaaataggaagacaa	181	FAM	50
			PKP2Marker_1R	gggatacagtgtgtagcaattta		-	
	PKP2-M2	TG	PKP2Marker_2F	ccaattcctgggctcaatag	341	HEX	*a*
			PKP2Marker_2R	tcctcagacatacaggcagaag		-	
ARVC-11 (DSC2)	DSC2-M1	CA	DSC2Marker_1F	cattccagggccgtagataa	286	HEX	*a*
			DSC2Marker_1R	gactgacagcagggacatca		-	
	DSC2-M2	GT	DSC2Marker_2F	tcttggatcctgctctgtga	291	FAM	*a*
			DSC2Marker_2R	gacaaacatacagctgccaaa		-	

Note: *DSP* = Desmoplakin; *PKP2* = Plakophilin-2; *DSC2* = Desmocollin-2; F = forward primer; R = reverse primer; Ta = Annealing temperature; *a* = Optimised by Michelle Parker

Table B-3: Primer sequences for mutation screening

Name of primer	Primer sequence (5'–3')	PCR product size (in bp)	Ta (in ^{o}C)
C10orf97Exon1F	aagctacttccaccgcttt	570	57
C10orf97Exon1R	gtgtctagaaagtccagcttca		
C10orf97Exon2F	aaaggttccttgatccacac	470	57
C10orf97Exon2R	aagtcagggaacgcctatta		
C10orf97Exon3F	gtgttctacagggctttcact	487	56
C10orf97Exon3R	cccactgatacttgactttctc		
C10orf97Exon4F	catcattaacccattcttactattg	583	56
C10orf97Exon4R	gcatctgaaactgaagcatatc		
C10orf97Exon5F	actgccgtagcatagctgaa	460	56
C10orf97Exon5R	caaattaatcctgaagcaatgg		
C10orf97Exon6F	cctgccaagtaggtgttctt	516	56
C10orf97Exon6R	taagcaacgagtgagtgctg		
C10orf97Exon7F	catttcaaatctgagttacttgtgg	384	54
C10orf97Exon7R	tcatacatcaagcaaacaatgc		
C10orf97Exon8F	ccactgtgctgatatatttggtg	453	54
C10orf97Exon8R	ctggctccaagattaatggtct		
C10orf97Exon9F	acccaccacaaacctgatcta	395	59
C10orf97Exon9R	gaagagagcggtactgaacca		
C10orf97Exon10F	aagtgctcgggcacaata	321	59
C10orf97Exon10R	gtgggaatagactgaagtttgc		
C10orf97Exon11F	atgtcattatgtctgcctaccc	364	57
C10orf97Exon11R	cacagaaacccaactcaaca		

C10orf97Exon12F	agggtctctaatagcaataggc	596	54
C10orf97Exon12R	ctagcaccatctacacagaacc		
C10orf97Exon13F	ccgaatgtgttgtagtaccagt	364	59
C10orf97Exon13R	gcactgcgttaccgtattc		
C10orf97Exon14F	cttcagggctttctggtagt	500	56
C10orf97Exon14R	caagggtaacgatctgtatgg		
C10orf97Exon15aF	actttcacattaatcagtaagcaga	641	54
C10orf97Exon15aR	aaatgctgtagtattgcgtttg		
C10orf97Exon15bF	agctttcatcaaagttcatcac	675	57
C10orf97Exon15bR	gctctcagggttattgggta		
HSPA14Exon1F	gtgcgcactgtgcagttt	381	*a*
HSPA14Exon1R	gctctcaagcctgtggatg		
HSPA14Exon2+3F	tttccacacagatggcaaag	452	*a*
HSPA14Exon2+3R	gcaatgaagaaacaaactatgctg		
HSPA14Exon4+5F	tgttcagtgatattgtgaaatacag	494	*a*
HSPA14Exon4+5R	gtgcaatcactctgcttctc		
HSPA14Exon6F	ggtgggcaagcaagtgtatt	389	*a*
HSPA14Exon6R	gcctcttggcatttcatca		
HSPA14Exon7F	tgtgaagaatatgggctctgaa	447	*a*
HSPA14Exon7R	ttccatccagaaaggtttgg		
HSPA14Exon8F	aaagggattaaactaagtgacttct	382	*a*
HSPA14Exon8R	tatccagtttccaagcaatg		
HSPA14Exon9F	cggcctgattcactaattt	377	*a*
HSPA14Exon9R	ggttgataaatctcaaacaaaca		
HSPA14Exon10F	tgatatttgtttgattgtactgagc	279	*a*

HSPA14Exon10R	aaatattctttcagtacccagagc		
HSPA14Exon11F	tgtaataaatggtgagatttcctc	598	*a*
HSPA14Exon11R	ttaagacctgtcagtatacaaggtt		
HSPA14Exon12F	tctctttgcctagtgtcctg	439	*a*
HSPA14Exon12R	ttcagaaatgagccttcataa		
HSPA14Exon13F	tcagtgataaataagctttggtcct	371	*a*
HSPA14Exon13R	ttcaagatgcaaatgacttgagat		
HSPA14Exon14F	cttgtttactgaaagccctcaag	547	*a*
HSPA14Exon14R	gcatcttactttctatgtggtcctg		
NMT2Exon1F	tggggagggacgaagagg	344	see Table C-3
NMT2Exon1R2	tcgtccagttccaggctc		
NMT2Exon1F2	ggaggcactagagggagc	283	see Table C-6
NMT2Exon1R	gccgaagaacccccagtc		
NMT2Exon2F	ttcctcacctacatcctaacacag	399	*b*
NMT2Exon2R	tggcattcaacacacagtca		
NMT2Exon3F	cgtgcgtctgctcaagttta	279	see Table C-4
NMT2Exon3R	tgttgttgttgttgttcatcca		
NMT2Exon4+5+6F	tcactgaccactttggatgg	732	58
NMT2Exon4+5+6R	gcttggcctgcattcagta		
NMT2Exon7F	agcctgcattcttattcttgg	399	51
NMT2Exon7R	aggttctagtctccatctctgct		
NMT2Exon8F	gtaacagccaagctgagacaa	297	*a*
NMT2Exon8R	ttcctcagacacaccagtcag		
NMT2Exon9F	ggctctgcagctctaactagc	489	56
NMT2Exon9R	tttcacttaagcatgtccttgat		

NMT2Exon10F	tgccgagaaagaagtgtgact	400	*a*
NMT2Exon10R	tcgcataaattccacctgtgt		
NMT2Exon11F	tcacggacgttaaatcctatt	448	*a*
NMT2Exon11R	gctgagatcatgccattg		
NMT2Exon12F	caggtgtgagctgagcctta	206	54
NMT2Exon12R	ttcttctttggaatggcagtt		
FAM107BExon1F	gcactaccctaccctgaagc	820	*a*
FAM107BExon1R	gctgaaacattcgccaca		
FAM107BExon2F	aaacctcagttctttaaggtgtgg	365	*a*
FAM107BExon2R	cagatgtcgtgcctggaata		
FAM107BExon3F	agcccagttacattccaagc	477	*a*
FAM107BExon3R	gaaagccttaacaaagtgcaaga		
FAM107BExon4F	caccatgttgcctgctgtag	390	*a*
FAM107BExon4R	ctggcacgtctggtaaatcc		
FAM107BExon5F	ggaagtaaagttgtcagggacttt	369	*a*
FAM107BExon5R	atccacccagatcgtaggaa		
PTERExon1F	tacgaagaggtagggcttgg	382	60
PTERExon1R	ctctttgcaacgtccctca		
PTERExon2aF	cgcctctccttaagtgcttg	388	60
PTERExon2aR	aggcagcagactgattgagc		
PTERExon2bF	ttgacttccaccaatgcaga	381	50
PTERExon2bR	ggatgtgtgaaaccgcaaat		
PTERExon3F	tgatttgccttcttacatgggta	691	60
PTERExon3R	tagcatgcccagctacctct		
PTERExon4F	ttggatttagtgcaacttgttgat	565	53

PTERExon4R	tgcgtgattagtattgcattga		
PTERExon5F	tcacatcattaagagccatttca	397	53
PTERExon5R	gttcggcctgcttactgttt		
PTERExon6F	tttacgacaagcacaataggtt	400	53
PTERExon6R	aactgattaatatctcacagtggac		
RPP38Exon1F	taacgaaacctgcactgctc	621	see Table C-8
RPP38Exon1R	acggtattggtggtgactcc		
RPP38Exon2F	ggacatcgggtcttgcttac	294	60 (Table C-2)
RPP38Exon2R	cccatgagttcgagaccact		
RPP38Exon3aF	gtcatgcagttgtcgctaaca	540	59
RPP38Exon3aR	ggcagggacacttctgctta		
RPP38Exon3bF	gccatgatcacctcacactt	584	59
RPP38Exon3bR	tctgtgccatagttacaaagagtc		
ARMETL1Exon1F	cccttcctatcccttccttg	393	55
ARMETL1Exon1R	cctgggcggaatgttattta		
ARMETL1Exon2F	ttaaatgtggttagtgggtca	300	55
ARMETL1Exon2R	gtaagtaggaggacttcacgtct		
ARMETL1Exon3F	caccatgtccagcctacctc	394	55
ARMETL1Exon3R	taccattcatcagccaagca		
ARMETL1Exon4F	cttccctgccttggatcata	400	55
ARMETL1Exon4R	tgatgcattcccagttatcct		
C10orf38Exon1F	gcggagccagcgacatgg	478	*a*
C10orf38Exon1R	gcggcgacagcagctgga		
C10orf38Exon2F	ccacttcccatcacagtcatac	400	60
C10orf38Exon2R	gcccaagctggtttcaga		

C10orf38Exon3F	actgcaatacacttgtagcattt	399	55
C10orf38Exon3R	tgatgggttaatgtgtttgat		
C10orf38Exon4F	aggcgtgctgaaaccttaga	372	*a*
C10orf38Exon4R	gcaccttcagtgcctttctc		
C10orf38Exon5F	ccctggcatataggaggtaatg	379	60
C10orf38Exon5R	ggaaaggagacttccatgagc		
C10orf38Exon6F	gacactttggccatcttggt	353	60
C10orf38Exon6R	tcgccctgctgaagttattt		
C10orf38Exon7F	gcacaagacactgcgttctc	287	60
C10orf38Exon7R	cccacctccatattaatgcttc		
C10orf38Exon8aF	tggctgctcctgacttgtt	500	60
C10orf38Exon8aR	gtcccacttggagtgaggtt		
C10orf38Exon8bF	cacctcccaggaatttagctc	600	60
C10orf38Exon8bR	ttggacagctcagcctgtagt		
C10orf38Exon8cF	gggaccactcctatgtcagc	562	60
C10orf38Exon8cR	gctcgaggctctttctgct		
C10orf38Exon8dF	gggagatgctttgtctctgc	489	60
C10orf38Exon8dR	ttcctccacgaacgggta		
DCLRE1CExon1F	ctagaacccgaccggatgctc	355	*a*
DCLRE1CExon1R	cctcctgtcctctctccagc		
DCLRE1CExon2F	ctgaggaggattcttcacttgcag	379	*a*
DCLRE1CExon2R	gtctgggatattggtactgtcc		
DCLRE1CExon3F	gtgagccacacagatgcaag	358	*a*
DCLRE1CExon3R	cactgatgtttgatacctgggctgg		
DCLRE1CExon4F	gttggtcagactggtctcaaatg	264	*a*

DCLRE1CExon4R	caatctacctctaaaaatacttcccac		
DCLRE1CExon5aF	cttctcattctctgtcgtgtgtg	365	*a*
DCLRE1CExon5aR	gcgggagaccagcactattgaggc		
DCLRE1CExon5bF	ggacagaccgaaacccacttctg	369	*a*
DCLRE1CExon5bR	cctatttcttgccctgattctgc		
DCLRE1CExon6F	accaaatctgcctcaagacaggta	358	*a*
DCLRE1CExon6R	ctgaggggtggagcatctgac		
DCLRE1CExon7F	gtgaaggagggctttgttgagc	447	*a*
DCLRE1CExon7R	caaatctgatcttcacctcagtc		
DCLRE1CExon8F	gcaggaagcatttagtatgcagctc	282	*a*
DCLRE1CExon8R	agaatcgcttgaactctgggcgac		
DCLRE1CExon9F	gtggtggcacaaacccatgatc	368	*a*
DCLRE1CExon9R	acagttaggatatgacctgtcacc		
DCLRE1CExon10F	ggtgacaggtcatatcctaactgt	489	*a*
DCLRE1CExon10R	tggtgaaatgtcgtctgtgct		
DCLRE1CExon11F	acatgctggcggaataccaga	415	*a*
DCLRE1CExon11R	cagttccatgaccttgagctaacaac		
DCLRE1CExon12F	gaagagtccagcccagttgttttg	436	*a*
DCLRE1CExon12R	cttgggcttaggctagactgtgc		
DCLRE1CExon13F	ggtgatttcgacacttgctgctg	257	*a*
DCLRE1CExon13R	gtcaactaccaaggctgcagaacac		
DCLRE1CExon14F	cagcattgagaaccacagccataag	276	*a*
DCLRE1CExon14R	ctcctttgtgtcctagccaagagc		
DCLRE1CExon15F	gtagttgtgtgatttcccttatggtc	317	*a*
DCLRE1CExon15R	caagctccctggaaagcagtg		

DCLRE1CExon16aF	gaagaagagtggcctccctat	399	*a*
DCLRE1CExon16aR	gcaggtgaagtacagagcccagatc		
DCLRE1CExon16bF	gctgcagagcagagtgtatgca	370	*a*
DCLRE1CExon16bR	tctgtagtcagctttgtccaagg		
DCLRE1CExon16cF	ctgggacagccaatctgatactgt	329	*a*
DCLRE1CExon16cR	cagcttctggagttgagggaacttc		
DCLRE1CExon16dF	tagcacaaatgcagattcccagagc	328	*a*
DCLRE1CExon16dR	caagccattgcccacctcaaagt		
C1QL3Exon1aF	ggagtattcagaatctcactgtt	366	*a*
C1QL3Exon1aR	cagcggcggcggcggcgggcacg		
C1QL3Exon1bF	cagaattctgccttcagtctc	365	*a*
C1QL3Exon1bR	cgccgccgcctgcgatcgac		
C1QL3Exon1cF	gaaagttggtgcgaacaggc	306	*a*
C1QL3Exon1cR	cggccgctgctgccgactcctgc		
C1QL3Exon1dF	ggacttcggctcaagtcac	360	*a*
C1QL3Exon1dR	ggtcgcagaccatgcggcaggtgc		
C1QL3Exon1eF	ctgctgctggtgatcctcatc	377	*a*
C1QL3Exon1eR	gcgatcttgggcaccgtg		
C1QL3Exon1fF	ctgcccaccttcatccag	330	*a*
C1QL3Exon1fR	ggtgaacttgccggtggtg		
C1QL3Exon1gF	cagcatgaaggctacgaggtg	274	*a*
C1QL3Exon1gR	ggctcccacccccatttc		
C1QL3Exon2aF	cttgagaattctcatttacgc	304	*a*
C1QL3Exon2aR	ggatccttgattcactgacg		

C1QL3Exon2bF	ctgagtttgaacactggattcg	335	*a*
C1QL3Exon2bR	gcagcaagatcacagtacac		
C1QL3Exon2cF	gtgaatgaagtcctatatagatc	375	*a*
C1QL3Exon2cR	gatcataggcttgggaattc		
C1QL3Exon2dF	ctgggaatgattcaacttcaagc	300	*a*
C1QL3Exon2dR	cggtttgatttgatactaaag		
SUV39H2Exon1F	agctgagctatcccgtcaga	466	*a*
SUV39H2Exon1R	ctcgaaccaccacaaaggag		
SUV39H2Exon2aF	tgttaggaacctgtgactacatg	502	*a*
SUV39H2Exon2aR	gctggtttgtattcgttaatgt		
SUV39H2Exon2bF	ctgcagagatggcaagatga	485	*a*
SUV39H2Exon2bR	aaccttgcatgtccatcattt		
SUV39H2Exon3F	gtagatgaatgcttctttggttcag	290	*a*
SUV39H2Exon3R	aaagggaattattctttagctatgg		
SUV39H2Exon4F	catgtagaagggcttgtttacacc	266	*a*
SUV39H2Exon4R	ctaaggtctgttcctggtctgatttg		
SUV39H2Exon5F	gtcttgaccatgtaagtaccctct	385	*a*
SUV39H2Exon5R	cagaggcatcggtaatacatttg		
MEIG1Exon1F	gctcctacctgggacctaca	462	*b*
MEIG1Exon1R	gatctccctggcaaacctc		
MEIG1Exon2F	ttgctctgtcacccagctt	618	*b*
MEIG1Exon2R	ttggagtcgaatttccttcag		
MEIG1Exon3F	acagcataaggctgcaatgt	692	*b*
MEIG1Exon3R	tggtagtgcacacctgtaattct		

C10orf111Exon1F	atggcgaatccgtggtca	395	*b*
C10orf111Exon1R	tataggctggcctccgaaca		
C10orf111Exon2F	ctttggaatcttagccctctg	698	*b*
C10orf111Exon2R	aatcacctgaacctgtgagg		
OLAHExon1F	acagcccatgtcattcttcc	243	*b*
OLAHExon1R	gataatccaactataggctgtgcat		
OLAHExon2F	ttctcctttgtcatcaccactg	386	*b*
OLAHExon2R	cacctggctgagaaacctct		
OLAHExon3F	tttgccctttcttattgga	380	*b*
OLAHExon3R	accaggaaggaaagtcttga		
OLAHExon4F	caggctggagtacaatagca	398	*b*
OLAHExon4R	aaatttaaagcttctgtgagtcag		
OLAHExon5F	ttgctggacttaactgttatcattc	400	*b*
OLAHExon5R	ggccataattacctaacatgctg		
OLAHExon6F	tgttgtgtttatgatactgttcctt	291	*b*
OLAHExon6R	tggaccagaattcaccagat		
OLAHExon7F	ttcccagccttggtttaatg	374	*b*
OLAHExon7R	tgtcttcccacataagcaacc		
OLAHExon8F	ccaagtttcattgacattaggaaca	296	*b*
OLAHExon8R	gctgagattgcgccattg		
OLAHExon9F	cgtacttcagggtgtcagctc	362	*b*
OLAHExon9R	acgaatctccctgacagtagaaa		
ACBD7Exon1F	tgtcttcgcctttccatttg	299	*b*
ACBD7Exon1R	ggagccccacttaagcactc		
ACBD7Exon2+3+4F	ttggaaccttcctcttcctg	680	*b*

ACBD7Exon2+3+4R	aacagtatgcctctccctctaaa		

Note: F = forward primer; R = reverse primer; Ta = Annealing temperature; *a* = Optimised by Gasna Shaboodien; *b* = Optimised by the Central Analytical Facility

Appendix C: Methods

C.1. Exclusion Mapping

C.1.1. PCR Optimisation

PCR was performed in 0.2ml tubes (Whitehead Scientific) in a total of 25µl containing 50ng of DNA, 0.4µM of each primer, 0.2mM of each of the 4 dNTPs (*Bioline,* Celtic Molecular Diagnostics), 1x PCR Buffer (*Promega* Colorless GoTaq^R Flexi Buffer; pH 8.5, Whitehead Scientific), 1.5mM of $MgCl_2$ (*Promega*, Whitehead Scientific), 0.5 unit of *Taq* polymerase (*Promega* GoTaq^R DNA Polymerase, Whitehead Scientific), made up with dH_2O. A total of eight tubes were set up each with the same DNA sample but a different annealing temperature. Two negative controls, where dH_2O replaced DNA, were included at the lowest annealing temperature and the highest annealing temperature to rule out contamination. The PCR cycling conditions are given in Table C-1.

Table C-1: PCR cycling conditions for microsatellite marker optimisation

Step	Temperature	Time	No. of cycles
Initial Denaturation	94^{o}C	5 minutes	1
Denaturation	94^{o}C	30 seconds	30
Annealing	44-58^{o}C*	30 seconds	
Extension	72^{o}C	30 seconds	
Final Extension	72^{o}C	7 minutes	1

*Each of the eight tubes had a different annealing temperature ranging from 44-58^{o}C.

C.1.2. Genotyping

A volume of 8µl of Hi-Di™ Formamide (*Applied Biosystems*) and 0.25µl to 0.5µl of GeneScan™ 500 ROX™ Size Standard (*Applied Biosystems*) was aliquoted into wells of a 96-well plate. At least 2µl of neat or diluted PCR product was added to the wells. The samples were denatured at 95^{o}C for 5 minutes using a thermal cycler and then cooled immediately on ice. The plate was loaded onto the ABI Prism® 3100 Genetic Analyzer (*Applied Biosystems*) and run with filter set D.

C.2. Mutation Screening by Sequencing

The primers for *MEIG1*, *C10orf111*, *OLAH* and *ACBD7* and DNA for two affected and two unaffected family members were sent to the Central Analytical Facility, Stellenbosch University for PCR optimisation, amplification of patient DNA and sequencing.

C.2.1. PCR Optimisation

PCR was performed in 0.2ml tubes (Whitehead Scientific) in a total of 25µl containing 100ng of DNA, 0.4µM of each primer, 0.2mM of each of the 4 dNTPs (*Bioline*, Celtic Molecular Diagnostics), 1x PCR Buffer (*Promega* Colorless GoTaq® Reaction Buffer, Whitehead Scientific; pH 8.5., 1.5mM $MgCl_2$), 0.5 unit of *Taq* polymerase (*Promega* GoTaq® DNA Polymerase, Whitehead Scientific), made up with dH_2O. A total of 12 tubes were set up each with the same DNA sample but a different annealing temperature. The PCR cycling conditions are given in Table C-2.

Table C-2: PCR cycling conditions for optimisation

Step	Temperature	Time	No. of cycles
Initial Denaturation	95°C	5 minutes	1
Denaturation	95°C	30 seconds	30
Annealing	50-60°C*	30 seconds	
Extension	72°C	30 seconds	
Final Extension	72°C	7 minutes	1

*Each of the 12 tubes had a different annealing temperature ranging from 50-60°C.

C.2.2. Optimised PCR Conditions

PCR of *NMT2* exon 1 amplicon 1, exon 3, 4+5+6, 7 and 12 required further optimisation. Several parameters were tested and the final, optimised conditions for these exons were as follows:

***NMT2* exon 1 amplicon 1**

PCR was performed in 0.2ml tubes (Whitehead Scientific) in a total of 25µl containing 100ng of DNA, 0.8µM of each primer (NMT2Exon1F & R), 0.2mM of each of the 4 dNTPs (*Kapa Biosystems*), 1x PCR Buffer (KAPA2G Buffer A, *Kapa Biosystems*; 1.5mM $MgCl_2$), 1x KAPAEnhancer 1 (*Kapa Biosystems*), 0.5 unit of *Taq* polymerase (KAPA2G Robust DNA Polymerase, *Kapa Biosystems*), made up with dH_2O. The PCR cycling conditions are given in Table C-3.

Table C-3: PCR cycling conditions for *NMT2* exon 1 amplicon 1

Step	Temperature	Time	No. of cycles
Initial Denaturation	95°C	1 minute	1
Denaturation	95°C	30 seconds	30
Annealing	63°C	30 seconds	
Extension	72°C	30 seconds	
Final Extension	72°C	1 minute	1

***NMT2* exon 3**

PCR was set up as in PCR optimisation (section C.2.1.). The PCR cycling conditions are given in Table C-4.

Table C-4: PCR cycling conditions for *NMT2* exon 3

Step	Temperature	Time	No. of cycles
Initial Denaturation	95°C	5 minutes	1
Denaturation	95°C	30 seconds	35
Annealing	59°C	1 minute	
Extension	72°C	1 minute	
Final Extension	72°C	7 minutes	1

***NMT2* exon 4+5+6**

PCR was set up as in PCR optimisation (section C.2.1.) except the PCR Buffer used was PCR optimization buffer C (*Roche* PCR Optimization Kit; pH 8.3., 2.0mM $MgCl_2$).

***NMT2* exon 7**

PCR was set up as in PCR optimisation (section C.2.1.) except the PCR Buffer used was PCR optimization buffer D (*Roche* PCR Optimization Kit; pH 8.3., 2.5mM $MgCl_2$).

***NMT2* exon 12**

PCR was set up as in PCR optimisation (section C.2.1.) except an additional 0.5mM $MgCl_2$ was added.

C.2.3. Sequencing of *NMT2* exon 1 amplicon 1

Purification of PCR products

Purification of PCR products was performed in 0.2ml tubes (Whitehead Scientific) in a total of 20µl containing 10µl of PCR product, 1 unit of SAP (*Promega*, Whitehead Scientific), 2 units of Exo I (*New England Biolabs*, Laboratory Specialist Services) made up with dH_2O. The tubes were incubated in a Labnet MultiGene Thermal Cycler (Whitehead Scientific) for 60 minutes at 37°C followed by 15 minutes at 75°C.

Cycle sequencing

Cycle sequencing was performed in 0.2ml tubes in a total of 20µl containing 5µl purified template, 3.2µM NMT2Exon1R or NMT2Exon1R2, 8µl BigDye® Terminator v1.1 Ready Reaction Mix (*Applied Biosystems*), 2µl DMSO, made up with dH_2O. The thermal cycling was carried out on a Labnet MultiGene Thermal Cycler (Whitehead Scientific) and the cycling conditions are given in Table C-5.

Table C-5: Sequencing cycling conditions

Temperature	Time	No. of cycles
96°C	5 minutes	1
96°C	30 seconds	30
50°C	15 seconds	
60°C	4 minutes	

Purification of sequencing products and capillary electrophoresis

The sequencing products were sent to the Central Analytical Facility, Stellenbosch University for purification and capillary electrophoresis.

Analysis

The electropherograms were analysed by eye to identify heterozygous variants. To identify homozygous variants, the patients' sequences were aligned to the wild type sequence. BioEdit was used to align the sequences and the wild type sequence was obtained from the NCBI database.

C.3. Mutation Screening by HRM

C.3.1. PCR and HRM Optimisation

PCR and HRM were performed in 0.2ml tubes (Whitehead Scientific) in a total of 25µl containing 100ng of DNA, 0.4µM of each primer, 1x Reaction Mix (*Quantace* SensiMix HRM™, Celtic Molecular Diagnostics; contains reaction buffer, heat-activated *Taq* DNA Polymerase, dNTPs, 3mM $MgCl_2$ and stabilisers), 1µl of EvaGreen™ (*Quantace*, Celtic Molecular Diagnostics), made up with dH_2O. The PCR and HRM conditions are given in Table C-6.

Table C-6: PCR and HRM conditions

Step	Temperature	Time	No. of cycles
Hold step (to activate enzyme)	95°C	10 minutes	1
Denaturation	95°C	5 seconds	50
Annealing	55°C	10 seconds	
Extension	72°C	10 seconds	
HRM	72-95°C (0.1°C increments)	-	1

C.4. Mutation Screening by DHPLC

C.4.1. PCR Optimisation

PCR was performed in 0.2ml plates in a total of 50µl containing 100ng of DNA, 0.4µM of each primer (*Invitrogen*), 0.04mM of each of the 4 dNTPs (*Invitrogen*), 1x PCR Buffer (*Promega* Colorless Go Taq® Flexi Buffer, pH 8.5.), 1.5mM $MgCl_2$ (*Promega*), 1 unit of *Taq* polymerase (*Promega* GoTaq® DNA Polymerase), made up with dH_2O. A total of 12 tubes were set up each with the same DNA sample but a different annealing temperature. The PCR cycling conditions are given in Table C-7. The melting temperature (Tm) was calculated for each primer using the formula: Tm = 4(G + C) + 2(A + T). The annealing temperature (Ta) was then calculated by subtracting 5°C from the lowest melting temperature for the primer pair. The calculated annealing temperatures were used to determine an appropriate temperature gradient.

Table C-7: PCR cycling conditions for optimisation

Step	Temperature	Time	No. of cycles
Initial Denaturation	95°C	2 minutes	1
Denaturation	95°C	30 seconds	30
Annealing	50-60°C*	30 seconds	
Extension	72°C	1 minute	
Final Extension	72°C	5 minutes	1
Hold	10 °C	-	1

* Each of the 12 tubes had a different annealing temperature ranging from 50-60°C.

C.4.2. Optimised PCR Conditions for *RPP38* exon 1

PCR of *RPP38* exon 1 required further optimisation. The final, optimised conditions for this exon are as follows:

PCR was performed in 0.2ml tubes in a total of 50µl containing 100ng of DNA, 0.4µM of each primer (*Invitrogen*), 0.04mM of each of the 4 dNTPs (*Invitrogen*), 1x PCR Buffer (*QIAGEN*; 1.5mM $MgCl_2$), 1 unit of *Taq* polymerase (*QIAGEN*), 1x Q-Solution, made up with dH_2O. A total of 12 tubes were set up each with the same DNA sample but a different annealing temperature. The PCR cycling conditions are given in Table C-8.

Table C-8: PCR cycling conditions for *RPP38* exon 1

Step	Temperature	Time	No. of cycles
Initial Denaturation	94°C	3 minutes	1
Denaturation	94°C	30 seconds	11
Annealing	66°C*	30 seconds	
Extension	72°C	1 minute	
Denaturation	94°C	30 seconds	30
Annealing	55°C	30 seconds	
Extension	72°C	1 minute	
Final Extension	72°C	7 minutes	1
Hold	10 °C	-	1

*Ta decreases by 1°C every cycle

C.4.3. Sequencing

Purification of PCR products

A volume of 5µl of PCR product was mixed with 2µl of ExoSAP-IT (*Amersham Bioscience*). ExoSAP-IT contains Exo I and SAP. The samples were incubated on a thermal cycler (GeneAmp PCR System 9700) at 37°C for 15 minutes and then incubated at 80°C for 15 minutes.

Cycle sequencing

Cycle sequencing was performed in 0.2ml tubes in a total of 10µl containing 1µl purified template, 0.4µM forward or reverse primer, 1µl BigDye® Terminator v1.1 Ready Reaction Mix (*Applied Biosystems*), 1.5µl Sequencing Buffer (*Applied Biosystems*), made up with dH_2O. The thermal cycling was carried out on a GeneAmp^R PCR System 9700 and the cycling conditions are given in Table C-9.

Table C-9: Sequencing cycling conditions

Temperature	Time	No. of cycles
96°C	1 minute	1
96°C	10 seconds	25
50 °C	5 seconds	
60 °C	4 minutes	
10 °C	Hold	1

Purification of sequencing products

The Montage SEQ_{96} Sequencing Reaction Cleanup Kit (*Millipore*) was used to purify the sequencing products. A volume of 20µl of Injection Solution was added the sequencing product. The sequencing product was transferred to a SEQ plate well. The SEQ plate was placed on a vacuum manifold with ice packs and a vacuum (20-25" Hg) was applied for 3 minutes 40 seconds. The SEQ plate was removed from the manifold and the bottom of the SEQ plate was blotted on tissue paper to remove excess liquid. A volume of 25µl of Injection Solution was added to the well. The SEQ plate was returned to the vacuum manifold and a vacuum was applied for 4 minutes. The SEQ plate was removed from the manifold and the bottom of the SEQ plate was blotted on tissue paper. A volume of 20µl of Injection Solution was added to the well and the SEQ plate was put on a microplate shaker at speed 2.5/3 for 10 minutes. The purified sequencing product was transferred from the well into a clean tube.

Capillary electrophoresis

A volume of 5µl of purified sequencing product was added to 15µl HiDi Formamide (*Applied Biosystems*). The sample was denatured at 95°C for 3 minutes on a thermal cycler and then rapidly chilled by placing the sample on ice or in a -20 °C freezer for 15 minutes. Sequencing was performed on the ABI Prism^{R} 310 Genetic Analyzer (*Applied Biosystems*) and analysed using the Sequence Analysis version 5.2.0 software (*Applied Biosystems*).

Analysis

The electropherograms were analysed by eye to identify heterozygous variants. To identify homozygous variants, the patients' sequences were aligned to the wild type sequence. BioEdit was used to align the sequences and the wild type sequence was obtained from the NCBI database.

Appendix D: Gel Formulas

1% (w/v) Agarose gel with ethidium bromide (100ml)

1g Agarose (*SeaKem® LE Agarose*, Whitehead Scientific)

100ml 1x TBE buffer (Appendix E)

6µl Ethidium bromide (10mg/ml)

3% (w/v) MS-8 Agarose gel with ethidium bromide (300ml)

9g Agarose MS-8

300ml 1x TBE buffer (Appendix E)

18µl Ethidium bromide (10mg/ml)

Stir the TBE buffer with a stirrer bar while adding Agarose MS-8. Stir for 5 minutes. Heat the mixture in a microwave for a few minutes at a time. Add the ethidium bromide to the solution, mix and pour into a casting tray. Allow to solidify. To achieve the best resolution, store the gel at 4°C for 30 minutes before use.

2% (w/v) Agarose gel with ethidium bromide (100ml)

2g Agarose (*Vivantis*)

100ml 1x TBE buffer (*Invitrogen*) or 1x TAE buffer (*EuroClone*)

7µl Ethidium bromide (*Vivantis*; 10mg/ml)

Appendix E: Solutions and Buffers

10x Tris Borate EDTA (TBE) buffer (2L)

216g Tris

110g Boric acid

14.8g EDTA

Make up to 2L with dH_2O

Green GoTaq® Flexi Buffer

The buffer contains a blue and a yellow dye. In a 1% agarose gel the blue dye migrates at the same rate as a 3-5kb DNA fragment and the yellow dye migrates at a rate faster than primers (<50bp) (Promega Product Information).

Sucrose solution

0.4g Sucrose

1ml dH_2O

Agarose Loading Dye

0.125g Bromophenol blue

20g Sucrose

Make up to 50ml with dH_2O and ensure the pH is 8 or more basic

Appendix F: DNA Ladders

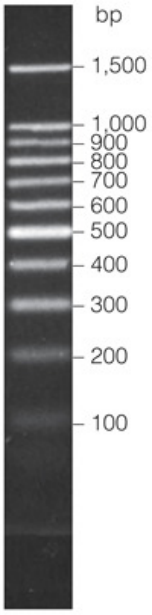


Figure F-1: 100bp DNA Ladder (*Promega*).

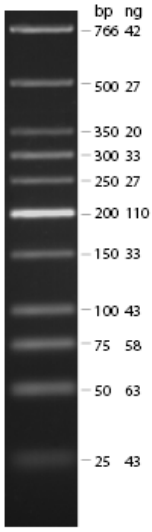


Figure F-2: Low Molecular Weight DNA Ladder (*New England Biolabs*).

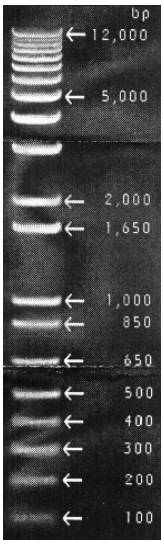


Figure F-3: 1 Kb Plus DNA Ladder (*Invitrogen*).

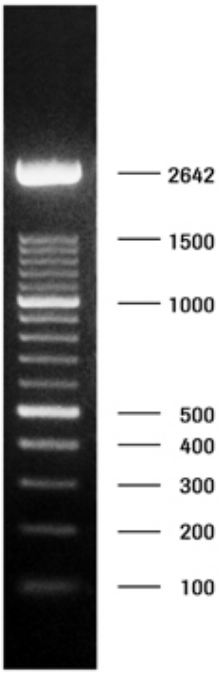


Figure F-4: DNA Molecular Weight Marker XIV (*Roche*).

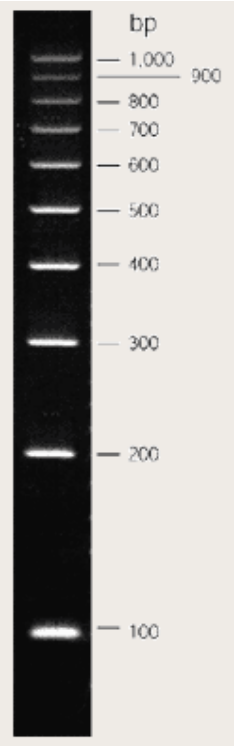


Figure F-5: Superladder-Low 100bp Ladder with ReddyRun™ (*ABgene*).

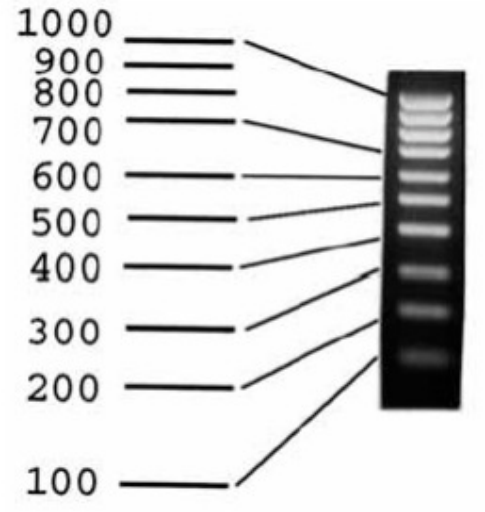


Figure F-6: 100bp DNA Ladder (*GENECRAFT*).

Appendix G: Candidate Genes in the Chromosome 4 Linkage Interval

Table G-1: Ensembl Gene IDs and HGNC Symbols for the 90 candidate genes in the chromosome 4 linkage interval

Ensembl Gene ID	HGNC Symbol
ENSG00000215367	TMED11P
ENSG00000249784	SCARNA22
ENSG00000249673	C4orf10
ENSG00000251166	OR4D12P
ENSG00000234386	OR7E84P
ENSG00000197701	ZNF595
ENSG00000250312	ZNF718
ENSG00000198155	ZNF732
ENSG00000198155	ZNF876P
ENSG00000131127	ZNF141
ENSG00000182903	ZNF721
ENSG00000182903	ABCA11P
ENSG00000174227	PIGG
ENSG00000133256	PDE6B
ENSG00000169020	ATP5I
ENSG00000215375	MYL5
ENSG00000169026	MFSD7
ENSG00000185619	PCGF3
ENSG00000168993	CPLX1
ENSG00000178950	GAK
ENSG00000127419	TMEM175
ENSG00000145214	DGKQ
ENSG00000145217	SLC26A1
ENSG00000127415	IDUA
ENSG00000127418	FGFRL1
ENSG00000178222	RNF212
ENSG00000159674	SPON2

ENSG00000159692	CTBP1
ENSG00000196810	C4orf42
ENSG00000090316	MAEA
ENSG00000163945	KIAA1530
ENSG00000179979	CRIPAK
ENSG00000235608	NKX1-1
ENSG00000174137	FAM53A
ENSG00000163950	SLBP
ENSG00000168936	TMEM129
ENSG00000013810	TACC3
ENSG00000068078	FGFR3
ENSG00000168924	LETM1
ENSG00000109685	WHSC1
ENSG00000185049	WHSC2
ENSG00000185818	NAT8L
ENSG00000243449	C4orf48
ENSG00000130997	POLN
ENSG00000214367	HAUS3
ENSG00000123933	MXD4
ENSG00000159733	ZFYVE28
ENSG00000063978	RNF4
ENSG00000125386	FAM193A
ENSG00000168884	TNIP2
ENSG00000087266	SH3BP2
ENSG00000087274	ADD1
ENSG00000109736	MFSD10
ENSG00000087269	NOP14
ENSG00000125388	GRK4
ENSG00000197386	HTT
ENSG00000188981	C4orf44
ENSG00000159788	RGS12
ENSG00000109758	HGFAC

ENSG00000175920	DOK7
ENSG00000163956	LRPAP1
ENSG00000184160	ADRA2C
ENSG00000163982	OTOP1
ENSG00000132406	TMEM128
ENSG00000145220	LYAR
ENSG00000168826	ZBTB49
ENSG00000168818	STX18
ENSG00000163132	MSX1
ENSG00000170891	CYTL1
ENSG00000152953	STK32B
ENSG00000082929	C4orf6
ENSG00000173040	EVC2
ENSG00000072840	EVC
ENSG00000072832	CRMP1
ENSG00000181215	C4orf50
ENSG00000152969	JAKMIP1
ENSG00000109501	WFS1
ENSG00000074211	PPP2R2C
ENSG00000013288	MAN2B2
ENSG00000179010	MRFAP1
ENSG00000163993	S100P
ENSG00000178988	MRFAP1L1
ENSG00000186222	CNO
ENSG00000170871	KIAA0232
ENSG00000132405	TBC1D14
ENSG00000173013	CCDC96
ENSG00000173011	TADA2B
ENSG00000109519	GRPEL1
ENSG00000207642	MIR571
ENSG00000216105	MIR943